



NATIONAL
CENTER *for* ANALYSIS of LONGITUDINAL DATA *in* EDUCATION RESEARCH

TRACKING EVERY STUDENT'S LEARNING EVERY YEAR

A program of research by the American Institutes for Research with Duke University, Northwestern University, Stanford University, University of Missouri-Columbia, University of Texas at Dallas, and University of Washington



*Has It Always Been
This Way?
Tracing the Evolution
of Teacher Quality
Gaps in U.S. Public
Schools*

Dan Goldhaber
Vanessa Quince
Roddy Theobald

Has It Always Been This Way? Tracing the Evolution of Teacher Quality Gaps in U.S. Public Schools

Dan Goldhaber
University of Washington Bothell/American Institutes for Research

Vanessa Quince
University of Washington, Seattle

Roddy Theobald
American Institutes for Research

Contents

Acknowledgements.....	ii
Abstracts.....	iii
1. Introduction.....	1
2. Prior Work.....	3
3. Data and Analytic Approach.....	8
4. Results.....	19
5. Discussion and Conclusion.....	29
References.....	33
Tables and Figures.....	38

Acknowledgements

The authors gratefully acknowledge financial support from the William T. Grant Foundation (grant #184925) and the National Center for Analysis of Longitudinal Data in Education Research (CALDER) funded through grant #R305C120008 to American Institutes for Research from the Institute of Education Sciences, U.S. Department of Education.

The research presented here is based in part on confidential data from the North Carolina Education Research Center (NCERDC) at Duke University, directed by Clara Muschkin and supported by the Spencer Foundation. We wish to acknowledge the North Carolina Department of Public Instruction for its role in collecting this information and making it available, as well as the Washington State Office of the Superintendent of Public Instruction (OSPI) for collecting the student-level data from Washington State and sharing it through data sharing agreement No. 2015DE-028.

We wish to thank James Cowan and Matt Kraft for helpful comments that improved the paper, as well as Joe Walch and Malcolm Wolff for research assistance.

CALDER working papers have not undergone final formal review and should not be cited or distributed without permission from the authors. They are intended to encourage discussion and suggestions for revision before final publication.

CALDER • American Institutes for Research
1000 Thomas Jefferson Street N.W., Washington, D.C. 20007
202-403-5796 • www.caldercenter.org

Has It Always Been This Way? Tracing the Evolution of Teacher Quality Gaps in U.S. Public Schools

Dan Goldhaber, Vanessa Quince, Roddy Theobald

CALDER Working Paper No. 171

December 2016

Abstract

There is mounting evidence of substantial “teacher quality gaps” (TQGs) between advantaged and disadvantaged students, but practically no empirical evidence about their history. We use longitudinal data on public school students, teachers, and schools from two states—North Carolina and Washington—to provide a descriptive history of the evolution of TQGs in these states. We find that TQGs exist in every year in each state and for all measures we consider of student disadvantage and teacher quality. But there is variation in the magnitudes and sources of TQGs over time, between the two states, and depending on the measure of student disadvantage and teacher quality.

1. INTRODUCTION

Income inequality has risen dramatically in the United States over the last 5 decades, and disparities in educational outcomes are likely a contributing factor to this trend. Many look to the public education system to close the achievement gaps that exist between advantaged and disadvantaged students when they start kindergarten (Lee & Burkham, 2002), but research on the extent to which schools are able to level the playing field is disheartening.¹ Although there is evidence that schooling interventions—the impact of teachers in particular (Chetty et al., 2014b; Rivkin et al., 2005; Rockoff, 2004)—*could* help close these gaps, they often do not because teacher quality itself is inequitably distributed across students.

Research has long shown that disadvantaged students are more likely to have “low quality” teachers—as measured by degrees, experience, and advanced credentials—than more advantaged students (e.g. Clotfelter et al., 2005; Kalogrides and Loeb, 2013; Lankford et al., 2002). More recent work also shows considerable inequity when teacher quality is measured using direct “value-added” measures of teacher effectiveness (e.g., Goldhaber et al., 2015; Isenberg et al., 2013; Sass et al., 2010), though one recent study suggests that the distribution of teacher effectiveness may be more equitable than this prior work suggests (Isenberg et al., 2016).

In response to mounting evidence of the importance of teachers and the existence of “teacher quality gaps” (TQGs), the federal government recently directed states to develop plans to reduce inequity in the distribution of teacher quality across public schools (Rich, 2014).

¹ For instance, there is evidence (Clotfelter et al., 2009) that achievement gaps between advantaged and disadvantaged students persist and often grow as students progress through the K–12 system.

Unfortunately, states are forced to develop these plans in what is close to an empirical vacuum about the *history* of these TQGs. That is, existing studies of TQGs represent snapshots in time and provide virtually no information about the sources of these gaps and how they have changed over time. This gap in our empirical knowledge is problematic, because the means of addressing teacher inequity depend fundamentally on the source of this inequity.

In this paper, we use longitudinal data on public school students, teachers, and schools from two “focal states”—North Carolina and Washington—to provide a descriptive history of the evolution and sources of TQGs. Data from these states include several different measures of teacher quality and student disadvantage that we use to calculate TQGs. Specifically, in each state we can measure teacher quality in terms of teacher experience, licensure test scores, and value-added estimates of effectiveness. Likewise, we can categorize students in each state as disadvantaged if they receive free/reduced lunch (FRL) or are from an underrepresented minority (URM) group.²

For each combination of teacher quality measure and student disadvantage measure, we calculate the corresponding TQG in each state and year as the average difference between disadvantaged and advantaged students in their exposure rates to low-quality teachers; for example, a novice teacher with fewer than 5 years of experience or a teacher in the lowest quartile of the distribution of licensure test scores or value-added.³ We then track the evolution of these TQGs in each state and investigate the extent to which each TQG is due to differences across districts, across schools within a district, and across classrooms within a school.⁴

² We define URM as American Indian, Black, or Hispanic.

³ We use the term exposure in a literal sense and do not intend any parallels to medical exposures.

⁴ We calculate classroom level TQG’s for years and grades in which student-teacher links are available.

We find that TQGs are not a new problem; in fact, disadvantaged students in both states are more likely to be exposed to low-quality teachers in every year of available data (going back to the late 1980s in one case) and under every definition we consider of student disadvantage (poor or minority) and teacher quality (experience, licensure test scores, and value-added estimates of effectiveness). But we also find some variation in the magnitudes of different TQGs over time and between the two states, as well as some important differences in the extent to which these TQGs are due to student and teacher sorting across districts, across schools within districts, and across classrooms within schools. This points to the importance of future work that disentangles the different factors that contribute to these TQGs.

The paper proceeds as follows. In Section 2, we review the prior literature that informs this study. We describe our data and analytic approach in Section 3, and then present our results in Section 4. We discuss implications for policy and directions for future research in Section 5.

2. PRIOR WORK

Our primary objective is to document the evolution of TQGs in two states using a number of different measures of student disadvantage and teacher quality. Although this question has not been explored in the existing empirical literature on inequities in U.S. public schools, we build on two subsets of this literature: the first provides different “snapshots” of TQGs at given points in time; and the other documents the evolution of *student achievement gaps* in public schools over time.

2.1. Snapshot Estimates of TQGs

This study builds most closely on prior work from Washington State (Goldhaber et al., 2015), one of the focal states in this study, which demonstrates that many different measures of teacher quality are inequitably distributed across various indicators of student disadvantage during the 2012 school year.⁵ The authors consider both input (teacher experience, licensure exam score) and output (value added) measures of teacher quality in their analysis and find that disadvantaged students (e.g., FRL and URM) are more likely to receive lower-quality teachers regardless of how teacher quality is measured. Furthermore, the authors demonstrate that student and teacher sorting across districts, across schools within districts, and across classrooms within schools all contribute to these TQGs.

Other studies have investigated TQGs using a subset of these measures of teacher quality and student disadvantage, although we again stress that none of these studies has considered the evolution of TQGs over time. For example, Lankford et al. (2002) find that lower-quality teachers (as measured by experience, degree, certification, and college of attendance) were more likely to teach in schools with higher numbers of low-performing minority students in the state of New York. Likewise, Clotfelter et al. (2005) use data from North Carolina (the other focal state in this study) and find that Black students are more likely to be in a classroom with a novice teacher than their White counterparts. This is crucial given the well-documented correlation between teacher experience and teacher effectiveness, particularly early in a teaching career (e.g., Rice, 2013; Rivkin et al., 2005; Rockoff, 2004). The authors disaggregate these results to classroom, school, and district effects and find significant effects at each level.

⁵ We use the convention that 2012 refers to the 2011–12 school year.

Kalogrides and Loeb (2013) focus on student and teacher assignments within schools themselves. They use data from three large school districts and find that classrooms with higher percentages of minority and low-income students were more likely to be assigned novice teachers. Kalogrides et al. (2013) delve further into this relationship and focus on one district from their previous study and find that less experienced teachers were more likely to be placed in classrooms with lower achieving students than their more experienced counterparts. These studies reinforce more qualitative evidence (e.g., Grissom et al., 2015; St. John, 2014), illustrating how experienced teachers typically have more discretion over their classroom assignments than less-experienced teachers.⁶

Although the preceding studies have focused primarily on teacher experience and other input proxies for teacher quality, researchers have also explored how output measures of teacher effectiveness (e.g., value-added estimates of teacher performance) are distributed across different groups. Sass et al. (2010) use student-level data from North Carolina and Florida and find that teachers in high-poverty schools on average, tend to have slightly lower value-added scores than those in other schools, but also that there is more variation in teacher value added in high-poverty schools. Hence, students in disadvantaged schools are considerably more likely to have a teacher in the bottom of the effectiveness distribution. For instance, teachers at the 10th percentile of the value-added distribution in disadvantaged schools in North Carolina are .04 standard deviations of student achievement less effective than teachers at the 10th percentile of the value-added distribution in advantaged schools in North Carolina.

⁶ Player (2010) also provides evidence that more *effective* teachers tend to be assigned to classrooms with higher-achieving and less-disadvantaged students.

Given the outsized effects of having a very ineffective teacher (Chetty et al., 2014b; Goldhaber and Startz, forthcoming; Hanushek, 1992), this suggests it is important to assess the tails of the value-added (and other measures of teacher quality) when considering TQGs.

A number of subsequent studies confirm the Sass et al. (2010) value-added TQG findings. Using data drawn from 10 school districts in seven states, Glazerman and Max (2012) find significant value-added TQGs at the middle school level (but not at the elementary level), and document substantial variation in the value-added distribution between the ten districts they consider. Isenberg et al. (2013) find significant and consistent value-added TQGs, and conclude that these differences are more attributable to teacher sorting across schools within these districts rather than teacher sorting across classrooms within schools. Most recently, Steele et al. (2015) find that schools within the highest quartile of minority students have teachers that are .11 standard deviations of student performance less effective than schools in the lowest quartile of the distribution of minority students.

The findings of large value-added TQGs are not, however, universal as several recent studies argue that the distribution of teacher effectiveness is relatively equitable (though all report positive and statistically significant TQGs). Chetty et al. (2014a) estimate that a \$10,000 increase in a student's family income is correlated with only a 0.001 increase in teacher value added. Mansfield (2015) reports that the average student in the bottom decile of a student background index is taught by a teacher at the 41st percentile of the value-added distribution, whereas the average student in the top decile of this index is taught by a teacher at the 57th percentile of value-added, and concludes that "teacher quality is fairly equitably distributed

both within and across high schools” (Mansfield, 2015, p. 751).⁷ Most recently, Isenberg et al. (2016) find only small gaps between the average value-added of teachers of FRL and non-FRL students.⁸

2.2 Evolution of Student Achievement Gaps

Our investigation of the evolution of TQGs builds on a small but influential literature documenting the evolution of *student* achievement gaps (i.e., the average difference in test achievement between advantaged and disadvantaged students) in public schools. Using data from 1962 through 1992, Hedges and Nowell (1998) show that the Black-White gap in student achievement decreased in the years following the Civil Rights era (1965–1972). For example, the gap in reading performance between Black and White students on the National Assessment of Educational Progress in 1994 is almost half of the comparable gap in 1975.

These findings on changes in the Black-White achievement gap are also echoed in a recent study by Reardon (2011) that assesses student achievement gaps over a 50-year period. Conversely, however, Reardon finds evidence that achievement gaps based on student poverty have increased over the same period and that the achievement gap by student poverty level is now twice as large as the racial achievement gap. This suggests that, although improvements have been made to address racial inequality, there is still much to be done in terms of addressing educational achievement gaps associated with family income.

⁷ Despite this characterization, the TQG between advantaged students reported by Mansfield (2015), .079, is in fact larger than the TQGs based on value-added from elsewhere in the literature (e.g., Goldhaber et al., 2015).

⁸ We note that Isenberg et al. (2016) only consider sorting within school districts, though much of the inequity in the distribution of value-added—at least in the one study that includes a state-wide analysis of TQGs based on value-added (Goldhaber et al., 2015)—is across school districts. We explicitly compare differences between the results in Isenberg et al. (2016) with the results in this paper, and other research discussed in this section, in Goldhaber et al. (2016c).

3. DATA AND ANALYTIC APPROACH

3.1. Context

North Carolina and Washington, the “focal states” in this study, provide interesting contrasts along several dimensions, and both have longitudinal data on students and teachers going back several decades. **Table 1** provides a comparison of these different state contexts and the data available in each state (and the years in which data are available). As shown in the statistics from the 2013 school year (the most recent year considered for both states) in Panel A, North Carolina has substantially more Black students and charter schools than Washington, but has less than half as many school districts despite a considerably larger overall student enrollment. As a result, the average district in North Carolina has more than three times as many students and more than twice as many schools as the average district in Washington. This difference is important in interpreting the extent of cross-district and within-district sorting in the two states, discussed in the next section.

Figure 1 illustrates how overall student demographics have changed in each state; the y-axis represents the proportion of URM students (Panels A and C)—defined as the proportion of students who are American Indian, Black, or Hispanic—or FRL students (Panels B and D), while the x-axis tracks these proportions over time. The general conclusion from these figures is that public school students within each state have become more racially and economically diverse over the time period studied. In North Carolina, for example, the percentage of URM students has increased from 33.6% in 1995 to 42.3% in 2014. The percentage of URM students in Washington has increased from 14.6% to 27.0% over the same time period.

Figure 2 illustrates the geographic distribution of student demographics across districts in each state; the shading within each figure represents the proportion of URM students (Panels A and C) or FRL students (Panels B and D) within each district. Figure 2 shows that URM and FRL students tend to be clustered within specific districts in both states (particularly URM students in Washington), and that districts with a high percentage of URM students tend to have a high percentage of FRL students; in fact, the correlation between the district-level percentages of URM students and the district-level percentages of FRL students is 0.67 in both Washington and North Carolina. Figure 2 also reinforces the statistics from Table 1 that North Carolina has a substantially more racially diverse student body and a somewhat higher-poverty student body than Washington's.

To illustrate how segregation across districts and schools in each state has changed over time, we calculate segregation indices for each year and state for URM students (Panels A and C) and FRL students (Panels B and D) and plot these indices in **Figure 3**. To calculate the segregation index (the y-axis in Figure 3) across districts in each state for a given measure of student disadvantaged (URM or FRL), we define D_l as the number of disadvantaged students in district l , ND_l as the number of nondisadvantaged students in district l , and compute the index of dissimilarity $\frac{1}{2} \sum_l \left| \frac{D_l}{\sum_l D_l} - \frac{ND_l}{\sum_l ND_l} \right|$. We follow a similar procedure to calculate the segregation index across schools in each state. In each case, this index can be interpreted as the proportion of students who would need to switch schools/districts to make the distribution of disadvantaged students perfectly equitable.

The fact that the school lines in each state lie above the district line shows that there is segregation both between and within districts.⁹ This is consistent with findings from other settings (e.g., Owens et al., 2016). We also see that the level of segregation across districts is more substantial in Washington than in North Carolina, especially in regards to the segregation of URM students where much of the school-level segregation is explained by segregation between districts.

3.2. Data Overview

For our primary analysis, we combine data from the focal states—provided by the North Carolina Education Research Data Center (NCERDC) and the Washington State Office of Superintendent of Public Instruction (OSPI)—with data from the National Center for Education Statistics (NCES) to create two different longitudinal datasets within North Carolina and Washington. First, we create a *school-assignment* dataset that enables us to consider the relationship between teacher characteristics and the aggregated student demographics of the teacher’s school. The advantage of these school-assignment datasets is that we can calculate TQGs across all grade levels and available years of data, but we cannot consider inequities in the *within-school* sorting of students and teachers with these datasets because we do not observe student or teacher classrooms assignments.

We also create a *student-assignment* dataset that takes advantage of the fact that student–teacher links are available in elementary grades in each state (since 1996 in North

⁹ For instance: were there no segregation between districts, the entire gap would be related to school level segregation within districts (and the district line would lie on the horizontal axis); were there no segregation across schools within districts, the entire gap would be related to district level segregation (and the two lines would lie on top of one another); and were there no gaps at all, both lines would be on the horizontal axis.

Carolina and since 2006 in Washington). The student-assignment data enable us to consider the relationship between teacher characteristics and the demographics of the students in a teacher's classroom within a given year. To facilitate cross-state comparisons, we focus on grade levels in which student-teacher matching is possible in each state across the greatest number of years (Grades 3–5), but we also investigate other grade levels within each state individually. While we can use the student-assignment datasets to calculate TQGs only in these grade levels and in years in which student-teacher links are available, the advantage of these datasets is that we can investigate the sorting of students and teachers across different classrooms within schools.

We describe the datasets for each state below, as well as the measures of teacher quality and student disadvantage that we employ throughout this analysis. Our primary results focus on the distribution of “low-quality” teachers, which we calculate from the teacher variables in both the school and student assignment datasets described below. For teacher experience, we focus on the distribution of novice teachers with 5 or fewer years of teaching experience, but also experiment with other definitions of “novice” (e.g., 2 or fewer years of experience).¹⁰ For licensure test scores and value-added estimates, we focus on the distribution of teachers who fall in the lowest quartile of the overall distribution of each teacher quality measure, and also experiment with other cut-points in these distributions (e.g., lowest decile), and further consider average teacher quality for advantaged and disadvantaged students in an extension.

¹⁰ The argument for using 5 or fewer years of experience as our measure is that studies of the student test achievement benefits of having a more experienced teacher (e.g., Rice, 2013; Rivkin et al., 2005; Rockoff, 2004) tend to show that much of the gains to teacher experience have been achieved by the time that teachers have five years of experience.

3.3. North Carolina Data

The school assignment dataset in North Carolina relies on teacher-level data from NCERDC going back to the 1995 school year, including information on the number of teachers, teacher position and salary (where experience is calculated) as well as teachers' average licensure test scores across reading, writing, and math scores on the state's licensure exam, the Praxis.¹¹ Our analysis of Praxis scores begins with the 2000 school year because this is the first year in which at least 1% of teachers in the state have a Praxis score, and we consider the average of each teacher's scores on the math, reading, and writing portions of the test from the *first* time each teacher took the test.¹²

For teachers who teach in tested grades and subjects,¹³ we include an estimate of the teacher's effectiveness calculated from the following value-added model (VAM) estimated for both math and reading:

$$Y_{ijst} = \beta_0 + \beta_1 Y_{i(t-1)} + \beta_2 X_{it} + \tau_{js} + \varepsilon_{ijst} \quad (1)$$

In (1), Y_{ijst} is the state test score for each student i with teacher j in subject s (math or reading) and year t , normalized within grade and year; $Y_{i(t-1)}$ is a vector of the student's scores the previous year in both math and reading, also normalized within grade and year; X_{it} is a vector of student attributes in year t (gender, race, FRL, English language learner status, gifted status, special education status, learning disability status); and τ_{js} is the VAM estimate that captures

¹¹ In North Carolina, teachers must also pass subject assessment tests, but those exams were not included in the analysis.

¹² Teachers may take licensure tests multiple times to get a passing score on all three tests, so we use the test scores from the first time each teacher took the Praxis (and follow a similar procedure with the WEST-B in Washington). This ensures that teachers taking the test for the fifth time, for example, are not judged as "comparable" to teachers who passed all three tests on the first attempt.

¹³ Eligible teachers include those who can be linked to students with a valid end-of-grade standardized achievement score in the current and prior year.

the contribution of teacher j to student test scores in subject s up to and including year t .¹⁴

Given that our interest is primarily in differential exposure to teachers at the tails of the value-added distribution (see Section 2), and teachers matched to a small number of students or with noisier value-added estimates are more likely to be in the tails of the distribution (e.g., Aaronson et al., 2007), for each year t we focus on pooled value-added estimates that consider all available years of data up to and including year t for each teacher. We also adjust all teacher effect estimates using empirical Bayes (EB) methods that shrink the estimates back to the grand mean of the value-added distribution proportionally to the standard error of each estimate. EB shrinkage does not account for the uncertainty in the grand mean, suggesting that the estimates may shrink too much under this procedure (McCaffrey et al., 2009), but this approach ensures that estimates in the tails of the distribution are not disproportionately estimates with large standard errors. We use the average math and reading value-added estimates for teachers who teach both subjects.¹⁵

The value-added specification in equation 1 is similar to the specification reported in Goldhaber et al. (2015). Isenberg et al. (2016) argue that one reason why the TQGs they report are smaller than the TQGs reported in Goldhaber et al. (2015) is because the primary value-added specification they rely on includes aggregated classroom covariates; i.e., the specification in equation 1 misattributes some of the impact of having disadvantaged classmates to

¹⁴ Because of computing limitations, we consider only up to 7 years of prior data in estimating these VAMs in North Carolina.

¹⁵ We also experiment with additional specifications of the model in equation 1, including a model that includes indicators for teacher experience level (so comparisons are made of students assigned to teachers with the same teaching experience).

differences in teacher quality between advantaged and disadvantaged classrooms. We experiment with models that include aggregated classroom covariates and find little evidence in either state that the inclusion of these variables in the VAM in equation 1 impacts the estimated TQGs at the elementary level (the grade level we consider in our primary analysis).¹⁶ This is not surprising given that Goldhaber et al. (2014) show that estimates of teacher value-added from models that do not include classroom characteristics are highly correlated with estimates from models that do ($r=0.99$).¹⁷

We merge the school assignment dataset to Public School Universe (PSU) data maintained by NCES. The PSU dataset includes school-level data about the percentage of students by race and ethnicity (linkable to North Carolina data since 1996) and the percentage of FRL students (linkable since 1999).¹⁸ We use the race and ethnicity variables to calculate the percentage of URM students in each teacher's school each year.¹⁹ The final school assignment dataset in North Carolina consists of 213,907 unique teachers and 1,554,901 teacher-year observations. We observe Praxis scores for 170,950 of these teacher-year observations, and value-added estimates for 148,312 teacher-year observations.

¹⁶ We provide details about the impacts of various VAM specifications on the estimates of TQGs in Goldhaber et al. (2016).

¹⁷ There is not a clear research consensus about whether VAMs that account for peer effects provide less biased estimates of teacher effects. For example, Kane et al. (2013) find that estimates from VAMs that do not account for peer effects have less forecast bias than estimates from VAMs that do, while Chetty et al. (2014b) find that estimates from a VAM with student-, classroom-, and school-level test score lags has less forecast bias than estimates from a model with just student-level lagged test scores.

¹⁸ We drop observations in the PSU data where total free-reduced lunch (totfrl) where less than or equal to zero. From 2011 and onward, we include other and mixed-race counts in the URM totals.

¹⁹ These school-level measures are highly correlated with school-level measures of academic performance. Specifically, the correlation between school percent URM and school average math performance is -0.62 in North Carolina and -0.48 in Washington; the correlation between school percent URM and school average reading performance is -0.70 in North Carolina and -0.61 in Washington; the correlation between school percent FRL and school average math performance is -0.77 in North Carolina and -0.66 in Washington; and the correlation between school percent FRL and school average reading performance is -0.83 in North Carolina and -0.74 in Washington.

We compile the North Carolina student assignment data from the NCERDC End-Of-Grade files and Masterbuild files. From 1995–2013, the dataset links students in Grades 3–5 with their classroom teachers.²⁰ The student-assignment dataset also includes student-level characteristics such as URM and FRL status, and we link these data to teacher experience, Praxis scores, and the VAM estimates described above.²¹ The final student assignment dataset in North Carolina consists of 37,374 unique teachers, 2,950,638 student-year observations, and 149,231 teacher-year observations. We observe Praxis scores for 16,798 of these teacher-year observations corresponding to 326,652 student-year observations, and value-added estimates for 92,601 teacher-year observations linked to 1,977,248 student-year observations.

3.4. WASHINGTON DATA

The school assignment dataset in Washington uses the state’s S-275 database, which contains information from OSPI’s personnel-reporting process and includes the school assignment of all certificated employees in the state in addition to a measure of each employee’s teaching experience in the state.²² Annual S-275 data are available from the 1984

²⁰ The End-of-Grade files do not explicitly link students with their classroom teachers, instead listing the employee who proctored the end-of-grade test. Consistent with other research using NCERDC data, we employ techniques to increase the reliability of our student–teacher matches. Specifically, we include only proctors who are full-time regular classroom teachers with assignments consistent with the grade level of their linked students, in self-contained classrooms of reasonable class size

²¹ One problem to note with the NCERDC End-Of-Course data is a lack of novice teachers in 2005. This includes a small number of teachers with 0 years of experience and, to a lesser extent, teachers with 1 year of experience. Our primary results using the student assignment data in North Carolina (e.g., in Figure 4 and Figure 6) simply skip this year, but we also use existing data from 2004–05 and 2005–06 to impute years of experience for missing teachers in 2005 using student-level characteristics such as race/ethnicity, gender, prior-year scores, FRL status, limited English proficiency status, special ed. status, and class size. Results using these imputed data are available upon request.

²² The S-275 contains the experience that teachers are credited with for pay purposes, which may not include out-of-state teaching, teaching in a private school, or substitute teaching.

school year through the 2015 school year, although we focus on years since 1988 because data in these years are linkable to the PSU data from NCES.

We link the S-275 to the same teacher quality measures describe above. First, we include teachers' test scores on the Washington Educator Skills Test—Basic, or WEST-B, a standardized test that all teachers must pass before entering a teaching education program. As with the Praxis in North Carolina, we consider the average WEST-B score across math, reading, and writing from the first time each teacher took the test. The WEST-B was required for entry into teacher education programs beginning in 2002, so we begin considering teacher WEST-B scores in the 2006 school year (the first year in which at least 1% of teachers in the state have a WEST-B score).

As in North Carolina, we merge the Washington school assignment dataset to the PSU data that provide school-level student counts by race and ethnicity (linkable to Washington data since 1988) and the percentage of FRL students (linkable since 2002). The final school assignment dataset in Washington consists of 100,875 unique teachers and 892,662 teacher-year observations. We observe WEST-B scores for 52,087 of these teacher-year observations, and value-added estimates from 40,009 teacher-year observations.

The student assignment dataset in Washington uses data from the state's Core Student Records System (CSRS) and Comprehensive Education Data and Research System (CEDARS), both maintained by OSPI. From 2006 through 2009, students in Grades 3–5 in the CSRS dataset can be linked to their classroom teacher by their proctor on the state exam.²³ Since 2010, the

²³The proctor of the state assessment was used as the teacher–student link for at least some of the data used for analysis. The *proctor* variable was not intended to be a link between students and their classroom teachers, so this link may not accurately identify those classroom teachers.

state’s CEDARS dataset allows all students to be linked to their classroom teachers through unique course IDs.²⁴ In our primary results, we focus on students in Grades 3–5 to facilitate comparisons across years and states, but we also consider student assignments in other grades since 2010 in extensions of our primary results. The student assignment dataset includes student-level FRL and URM variables, and we link these data to the same teacher variables (experience, WEST-B scores, and VAM estimates) described above. The final student assignment dataset in Washington consists of 17,772 unique teachers, 1,423,347 student-year observations, and 66,561 teacher-year observations. We observe WEST-B scores for 13,912 of these teacher-year observations corresponding to 274,070 student-year observations, and value-added estimates for 35,138 teacher-year observations linked to 800,252 student-year observations.

3.5. ANALYTIC APPROACH

Our methodology for calculating TQGs in each school year from student assignment data closely follows the approach of Clotfelter et al. (2005) and Goldhaber et al. (2015), so we present our approach to calculating TQGs from the school assignment datasets described above. First, for a given measure of teacher quality (experience, licensure test score, or VAM), let X_{klt} be the proportion of “low quality” teachers in school k , district l , and year t . Likewise, for a given measure of student disadvantage (URM or FRL), let D_{klt} be the number of disadvantaged students in school k , district l , and year t (and let ND_{klt} be the corresponding number of

²⁴ CEDARS data include fields designed to link students to their individual teachers, based on reported schedules. However, limitations of reporting standards and practices across the state may result in ambiguities or inaccuracies around these links.

nondisadvantaged students). The school-level exposure rate of disadvantaged students to low-quality teachers is calculated as the following weighted average:²⁵

$$E_D(\bar{X}_{klt}) = \frac{1}{\sum_k \sum_l D_{klt}} \sum_k \sum_l X_{klt} D_{klt} \quad (2)$$

$E_D(\bar{X}_{klt})$, which is a proportion bounded by zero and one, can be interpreted as a measure of the average school-level proportion of low-quality teachers for disadvantaged students in year t . Likewise, school-level exposure rate of nondisadvantaged students to low-quality teachers is calculated as a similar weighted average, representing the average school-level proportion of low-quality teachers for nondisadvantaged students in year t :

$$E_{ND}(\bar{X}_{klt}) = \frac{1}{\sum_k \sum_l ND_{klt}} \sum_k \sum_l X_{klt} ND_{klt} \quad (3)$$

For this measure of teacher quality and student advantage, the school-level TQG in year t is simply $E_D(\bar{X}_{klt}) - E_{ND}(\bar{X}_{klt})$, or the difference in the average school-level exposure rates to low-quality teachers between disadvantaged and advantaged students in year t . We can follow a similar procedure to calculate a corresponding district-level TQG in year t , $E_D(\bar{X}_{lt}) - E_{ND}(\bar{X}_{lt})$. Thus the portion of the school-level TQG that is due solely to the sorting of students and teachers across schools within the same district can be calculated as $(E_D(\bar{X}_{klt}) - E_{ND}(\bar{X}_{klt})) - (E_D(\bar{X}_{lt}) - E_{ND}(\bar{X}_{lt}))$.

²⁵ This methodology pools results across all grade levels (K-12). As discussed in Goldhaber et al. (2015), one concern with this approach is that students are less likely to receive FRL as they progress through the schooling system, so aggregating FRL results across grade levels may misattribute differences in teacher quality across grade levels to differences in teacher quality between different types of students. We therefore investigate some individual grade levels as part of the extensions in Section 4.5.

4. RESULTS

4.1. Long-Term Trends for Between- and Within-District TQGs

We first use the school assignment data described in Section 3 to calculate TQGs at the school and district level in both states, in every year of available data, and using every combination of student disadvantage measure (URM and FRL) and measure of low teacher quality (novice, bottom-quartile licensure test, bottom-quartile value-added). Because of the sheer number of TQG estimates, we present all results in a series of figures.²⁶

Figure 4 traces the evolution of TQGs in terms of student exposure to novice teachers (with five or fewer years of experience) in the school assignment data in each state for URM students (Panels A and C) and FRL students (Panels B and D).²⁷ For example, Panel A in Figure 4 shows the average school-level (black lines) and district-level (red lines) proportion of novice teachers for URM students (solid lines) and non-URM students (dashed lines) in North Carolina. In other words, the solid black line shows the evolution of school-level exposure to novice teachers for URM students ($E_D(\bar{X}_{klt})$), the dashed black line shows the evolution of school-level exposure to novice teachers for non-URM students ($E_{ND}(\bar{X}_{klt})$), the solid red line shows the evolution of district-level exposure to novice teachers for URM students ($E_D(\bar{X}_{lt})$), the dashed red line shows the evolution of district-level exposure to novice teachers for non-URM students ($E_{ND}(\bar{X}_{lt})$).

²⁶ We do not perform any statistical tests of these results because each set of results comes from a complete census of a given set of students and teachers (e.g., all students and teachers in the state for the experience analysis, students and teachers in tested grades and subjects for the VAM analysis, etc.), and results should only be generalized to that set of students and teachers.

²⁷ We investigate other definitions of novice teacher in Section 4.5.

The vertical distance between the red lines in each year represents the district-level TQG in terms of exposure to novice teachers in each year ($E_D(\bar{X}_{lt}) - E_{ND}(\bar{X}_{lt})$), while the vertical distance between the black lines represents the school-level TQG in that year ($E_D(\bar{X}_{klt}) - E_{ND}(\bar{X}_{klt})$). We plot the magnitudes of these TQGs over time in the bar plot at the bottom of each figure (refer to the right axis for the magnitudes). The red portion of each bar can be interpreted as the portion of the TQG that is due to student and teacher sorting across districts (i.e., “cross-district” sorting), while the black portion of each bar can be interpreted as the portion of the TQG that is due to student and teacher sorting across schools within the same district (i.e., “within-district” sorting).

Before discussing the details of these plots (and the plots derived from the student assignment data in Figures 8-10), we pause to note a fundamental conclusion from these figures. In every single year of observed data in each state, and across every combination of student disadvantage and teacher quality, the TQG is positive; i.e., disadvantaged students are more likely to be exposed to low-quality teachers. Though consistent with the existing literature discussed in Section 2, this drives home the reality that TQGs are pervasive and *not* a new phenomenon in either of these states. The remainder of this discussion focuses on the trends within each state for given measures of student disadvantage and teacher quality, as well as the differences between the two focal states; these differences emphasize the importance of looking beyond a single state in research like this.

Figure 4 shows that TQGs in North Carolina in terms of exposure to novice teachers are largely due to within-district sorting for FRL students (Panel B), but there is considerable sorting across districts in North Carolina for URM students (Panel A). A second conclusion from this

figure is that the TQGs in terms of exposure to novice teachers in Washington (Panel C) have changed considerably over time; for example, the gaps for URM students are largely due to within-district sorting in the late 1980s, but while the extent of within-district sorting has remained remarkably consistent over the subsequent decades, the growing gaps due to cross-district sorting have caused these TQGs to grow considerably since the 1980s. In fact, while URM students were only 10% more likely to be exposed to a novice teacher in 1988, they were 34% more likely to be exposed to a novice teacher by 2013. Finally, in both states (but particularly Washington), the TQGs in terms of exposure to novice teachers are larger for URM students than for FRL students.

Figure 5 shows the evolution of TQGs in terms of student exposure to teachers with low licensure test scores in the school assignment data in each state. This figure illustrates an important difference between the sources of TQGs in North Carolina and Washington; while these cross-district sorting and within-district sorting contribute approximately equally to these gaps in North Carolina, cross-district sorting is responsible for a far greater share of these TQGs in Washington. This may partially be due to the difference in segregation patterns shown in Figure 3, as much of the segregation of URM students (and, to a lesser extent, FRL students) in Washington is between different districts. It may also be due to the existence of larger districts in North Carolina (and thus more opportunity for within-district sorting); in the extreme, a state with only one district only has within-district sorting, so we would expect more within-district sorting when there are more schools in each district (as is the case in North Carolina relative to Washington). That said, we still find more within-district sorting in North Carolina when we limit the sample in each state to “large districts,” those with at least 10,000 students, as well as

when we restrict the Washington data to the 34 districts in the Puget Sound Education Service District (ESD) that are also some of the largest in the state of Washington.²⁸ This suggests that these patterns are not driven solely by the larger districts in North Carolina. That said, the magnitude of the TQGs with respect to licensure test scores are remarkably consistent for different measures of disadvantage, over time, and between the two states; disadvantaged students are between 5 and 10 percentage points more likely to be exposed to a teacher with a low licensure test scores than nondisadvantaged students.

Figure 6 illustrates a similar difference between the focus states—i.e., within-district sorting historically accounts for most of the TQGs in terms of exposure to low-value-added teachers in North Carolina, while cross-district sorting accounts for most of the analogous TQGs in Washington—but also illustrates that these TQGs can sometimes change considerably over time. For example, the low-value-added TQG for URM students in Washington is over twice as large in 2009 and 2010 than in 2012-2014. Overall, though, disadvantaged students (URM or FRL) are between 3 and 8 percentage points more likely to be exposed to a low-VAM teacher in any given year, and in either state, than non-disadvantaged students.

Given well-documented returns to early teacher experience, the patterns in Figure 6 may be partly driven by differential exposure to novice teachers (shown in Figure 4). However, we replicate these figures using estimates from VAMs that control for teacher experience and still find consistent (though smaller) TQGs over time, meaning that even between students

²⁸ The average district in the Puget Sound ESD has 11,539 students, which is comparable to the average district size in North Carolina (12,235 students) but about three times larger than the average district in Washington (3562 students).

assigned to a teacher of the same experience level, disadvantaged students are more likely to have a low-value-added teacher than advantaged students.²⁹

4.2. Long-Term Trends for Within-School TQGs

We next use the student assignment data from each state described in Section 3—that, importantly, only include students in Grades 3-5—to calculate TQGs at the district, school, and here, the *classroom* level over time in terms of exposure to novice teachers (**Figure 7**), teachers with low licensure test scores (**Figure 8**), and teachers with low value-added estimates (**Figure 9**). The blue lines in each plot represent the proportion of novice teachers in the classrooms of different types of students, while the blue portion of the bars at the bottom of each plot can be interpreted as the portion of the overall TQG that is due to the sorting of students and teachers across classrooms within the same school (i.e., “within-school” sorting). As the patterns within these figures are broadly consistent with the analogous figures derived from the school assignment data, we focus on conclusions from these figures that go beyond the conclusions derived from the school assignment data.

First, in both states, within-school sorting contributes a small but meaningful portion of the TQGs in terms of exposure to novice teachers (Figure 7) and low-value-added teachers (Figure 9), but less in terms of exposure to low licensure test teachers (Figure 8).³⁰ Second, there appears to be more within-school sorting, particularly in terms of low-value-added teachers (Figure 9), in North Carolina than in Washington. This holds even when we limit the data to schools in each state that have at least one novice teacher and at least one non-novice

²⁹ Results available from authors upon request.

³⁰ We note that Goldhaber et al. (2015) finds more evidence of within-school sorting in middle schools and high schools, which is not surprising given the prevalence of tracking at these grade levels.

teacher.³¹ Finally, the magnitudes of TQGs calculated from the student assignment data are quite similar to the TQGs from the school assignment data, suggesting that the overall magnitudes of TQGs in Figures 7-9 are broadly representative of the TQGs from other grade levels.

4.3. Cumulative Exposure to Low-Quality Teachers

The TQGs discussed in Sections 4.1 and 4.2 show that disadvantaged students are consistently more likely to be exposed to low-quality teachers than more advantaged students. While this clearly matters to students on an annual basis, these consistent TQGs also have the cumulative effect of exposing disadvantaged students to more low-quality teachers over time than advantaged students. To illustrate this cumulative effect, we take the exposure rates to low-quality teachers from the 2008 through 2013 school years from the student assignment data (Figures 7-9), assume that these exposure rates hold for all grades in elementary schools, and use a Poisson Binomial distribution to calculate the probability that first-graders in 2008 will be exposed to zero, one, up to six low-quality teachers by the time they reach sixth grade in 2013. We present these estimates as cumulative probabilities in **Figures 10-12**. In these figures, the point corresponding to a value a on the x-axis represents the estimated probability that a student will have a or fewer low-quality teachers; i.e., higher values on the y-axis represent *less* cumulative exposure to low-quality teachers.

The patterns across these figures are quite consistent across the different measures of teacher quality and student disadvantage. For example, Panel A of Figure 10 illustrates that URM students have only a 43% chance of having one or fewer novice teacher during their

³¹ Results available from authors on request.

elementary school years in North Carolina, compared to 57% for non-URM students, and that URM students in North Carolina are twice as likely to have three or more novice teachers during this time than non-URM students. This illustrates the extent to which URM and FRL students are consistently placed at a cumulative disadvantage in terms of exposure to low-quality teachers.

4.4 Heterogeneity in TQGs Across Districts

The results discussed to this point are pooled across districts in North Carolina and Washington, but prior work from these focal states (Clotfelter et al., 2005; Goldhaber et al., 2015) shows that there is considerable heterogeneity in TQGs between different districts within these states. The maps in **Figure 13** illustrate the TQGs with respect to novice teachers for URM students (Panels A and C) and FRL students (Panels B and D) for every district in each state, calculated from the most recent year of data in each state; districts with darker shading have higher TQGs with respect to novice teachers.³² The maps in Figure 13 illustrate that the clear trends in Figure 4 (illustrating that FRL and URM students are more likely to be exposed to novice teachers, on average, within both states) mask considerable heterogeneity across districts in terms of the inequitable exposure of URM and FRL students to novice teachers. That said, more districts have positive TQGs than negative TQGs within each map; for example, 70% of districts in Washington have a positive TQG between FRL and non-FRL students.

Interestingly, the correlations between these TQGs in Figure 13 and the corresponding student

³² We focus on the TQGs for novice teachers because we observe teacher experience for every teacher within each state. Note that the TQGs within each figure are constrained to be between -0.1 and 0.1—i.e., the small number (less than 1%) with TQGs outside this range are plotted as -0.1 or 0.1—to make differences within each figure more easily visible.

demographics are relatively weak ($|r| < 0.2$ in each case), suggesting that the overall diversity of a district is not highly predictive in either state of the extent of inequitable sorting within the district.

When we investigate district-level TQGs using other measures of teacher quality, a number of interesting patterns emerge.³³ First, while district-level TQGs are quite similar whether the measure of student disadvantage is FRL or URM ($r > 0.6$), there is more divergence depending on the measure of teacher quality considered. In Washington, for example, district-level TQGs according to licensure test scores are not significantly correlated with TQGs according to experience and value added, while district-level TQGs with respect to novice teachers and value added are significantly but only weakly correlated ($r = 0.16$).³⁴ While the correlation between the novice teacher and value-added TQGs is not surprising given the well-documented early-career returns to teacher experience, our broad conclusion is that districts that have large TQGs according to one measure of teacher quality may not have large TQGs according to other measures.

We further investigate the variability in district-level TQGs according to the same measures of teacher quality but within districts across time. We find that a district's prior year TQG is quite predictive of its TQG in the current year; for example, when we predict a district's TQG in 2013 as a function of its TQG in 2012, coefficients range from about 0.4 (for value added) to 0.7 (for novice teachers and licensure test scores). Some of the variation across time

³³ For this exploration, we use the student assignment data so all TQGs are comparable across measures of teacher quality (i.e., for students in grades 4 and 5), limit data to districts with at least 100 students in these grades, and (for the within-year correlations) focus on the most recent year of data within each state.

³⁴ The correlation is somewhat higher when we define novice teachers as teachers with two or fewer years of experience ($r = 0.21$).

could be due to changes in student assignments and teacher staffing across years within districts, while there is likely to be additional variability in the value-added TQGs due to changes in value-added estimates for individual teachers over time (e.g., Goldhaber and Hansen, 2013).

4.5. Robustness of Findings to Different Measures

We conclude by testing the robustness of our findings to different approaches to estimating TQGs, different measures of student disadvantage, and different definitions of low teacher quality. First, while the results discussed to this point focus exclusively on differences in the exposure rates to *low-quality* teachers between advantaged and disadvantaged students, much of the prior work discussed in Section 2 focuses on the differences in *average* teacher quality between advantaged and disadvantaged students. To place our results in the context of this broader literature, we present the evolution of TQGs in terms of average teacher quality in **Figures 14-19**.

These figures illustrate additional patterns that are not apparent from the figures that focus solely on the lower end of the distribution of each measure of teacher quality. For example, Panel C of Figure 14 demonstrates that URM students in Washington attended districts with *higher* average teacher experience in the late 1980s. However, this changed by the mid-1990s, and as of 2013 URM students attend districts with almost one fewer year of average teacher experience than non-URM students. Figure 15 reinforces the large gaps in average licensure test scores between teachers of advantaged and disadvantaged students in both North Carolina and Washington over time; for example, the gaps in Washington (Panels C and D) represent about 20 percent of a standard deviation of candidate performance on the WEST-B.

Finally, the gaps shown in Figure 19 are similar in magnitude to the gaps discussed in some of the earlier work exploring value-added TQGs (e.g., Goldhaber et al., 2015; Isenberg et al., 2013, 2016; Mansfield, 2015; Sass et al., 2011). At the elementary level the magnitude of the TQGs in North Carolina and Washington are consistently between .02 and .04 standard deviations of student achievement, regardless of the measure of student disadvantage that we consider. In the most recent year (2013), the magnitude of the gap implies that the average FRL elementary student in North Carolina is taught by a teacher at the 45th percentile of the value-added distribution, while the average non-FRL elementary student is taught by a teacher at the 55th percentile.³⁵

We also experiment with other measures of student disadvantage (e.g., students who receive free lunch instead of free or reduced priced lunch, Hispanic or Black students instead of URM, etc.) and definitions of low teacher quality (e.g., defining novice teachers as teachers with two or fewer years of experience, low licensure test teachers and low value added teachers as being in the lowest decile of the distribution, etc.) to ensure that our findings are robust to these different definitions. Our overall conclusion is that the patterns from Sections 4.1-4.4 are generally robust to these different definitions. There are, however, a few interesting departures from our main results. For example, we find that Black students in Washington were exposed to *fewer* novice teachers than White students in Washington prior to 1995, but the novice teacher gap between Black and White students has been positive for the past 20 years (the novice teacher gap between Hispanic and White students has been positive in all years of data).

³⁵ The comparable figures in Washington are 47th percentile for FRL students and 53rd percentile for non-FRL students.

It is also interesting to note that TQGs calculated from more restrictive definitions of low teacher quality (teachers with two or fewer years of experience and/or teachers in the lowest decile of the distributions of licensure test scores or value added) are larger in percentage terms than the TQGs reported in Section 4.1. This means that disadvantaged students are particularly likely to be exposed to teachers at the lower-tail of the effectiveness distribution in both states. This recalls findings from Sass et al. (2010) that the TQG between low-poverty and high-poverty schools is primarily driven by the presence of more low value-added teachers in high-poverty schools.

5. DISCUSSION AND CONCLUSIONS

The broadest conclusion from this analysis (already discussed in Section 4) is that TQGs are *not* a new phenomenon; in fact, disadvantaged students in both states were more likely to be exposed to low-quality teachers in every single year of available data and under every definition of student disadvantage and teacher quality. TQGs are therefore a persistent feature of public schools that only exacerbate well-documented achievement gaps between advantaged and disadvantaged students.

That said, a number of trends point to potential implications for policy. First, TQGs have historically been larger when student disadvantage is defined by race than by poverty level. For example, the difference in the exposure rate to novice teachers in both states between URM and non-URM students has been typically about twice as large as the corresponding difference between FRL and non-FRL students. Moreover, this “novice teacher gap” between URM and

non-URM students has grown considerably in each state, particularly in Washington. This suggests that, in contrast to evidence that gaps in student *performance* by race have been decreasing over the past several decades (e.g., Reardon, 2011), gaps in exposure to inexperienced teachers by student race have only grown over time. This is not surprising given evidence from the teacher labor market literature suggesting teacher preferences for schools with fewer minority students (e.g., Engel et al., 2013) that also tend to have stronger school organizational contexts (e.g., Kraft et al., 2015). Policies that incentivize teachers to work in high-minority schools—such as a bonus policy in North Carolina that considerably reduced the attrition of targeted teachers from high-minority schools in the 2000s (Clotfelter et al., 2008)—may be a fruitful avenue for policymakers looking to close these gaps. That said, the gaps by student poverty level are also educationally meaningful, particularly in light of recent evidence demonstrating the importance of school quality to intergenerational income mobility in the U.S. (Chetty et al., 2014c).

Second, there are important differences in the history of TQGs depending on the measure of teacher quality we consider. While TQGs by teacher experience have grown over time, corresponding gaps by teacher licensure test scores in both states have been quite consistent over the years of available data, and gaps by value-added estimates have varied considerably. This points to the importance of understanding the processes that contribute to these TQGs to explain this variation between teacher quality measures and over time.

Our investigation of the heterogeneity of TQGs across different districts and the extent to which each TQG is due to differences across districts, across schools within a district, and across classrooms within a school begins to point us in this direction. For example, while TQGs

by teacher experience in Washington in the late 1980s were primarily due to student and teacher sorting within districts, sorting across districts has been the more important contributor to TQGs in Washington for most of the past 20 years. Within-district sorting also contributes to TQGs far more in North Carolina than in Washington, across all measures of teacher quality we consider. This is important because, while prior work in Washington suggests that seniority transfer provisions in CBAs may be an important contributor to within-district inequities in teacher quality (Goldhaber et al., 2016b), districts in North Carolina are not bound by collectively-bargained personnel laws yet appear to have even more within-district inequity.

That said, the goal of our future research agenda is to examine the extent to which different *processes* in public schools contribute to TQGs and their evolution. Specifically, the TQGs described in this paper are the result of four different processes. First, changing student demographics in different classrooms, schools, and districts may be an important process that contributes to these gaps, particularly given recent evidence of increased racial segregation (Reardon and Owens, 2014) and income segregation (Owens et al., 2016) across schools and school districts. In other words, it's possible that changes in TQGs (e.g., the growing novice teacher gaps in Washington) are due in large part to growing disadvantaged student populations in districts that *already had* more novice teachers.

The other three processes have all been well-studied in the teacher labor market literature, but we do not know how important each process is in contributing to TQGs. Specifically, teachers in disadvantaged schools are far more likely to leave their school than teachers in more advantaged schools (Goldhaber et al., 2011; Hanushek et al., 2004; Scafidi et al., 2007), teachers who decide to transfer between schools tend to transfer into schools with

more advantaged students than the school they left (Clotfelter et al., 2011), and disadvantaged schools tend to hire far more inexperienced and under-qualified teachers than advantaged schools (Darling-Hammond, 2004). Future research into how these processes contribute to TQGs will inform the *process* that policymakers should seek to influence to close TQGs. For example, if patterns in teacher hiring explain most of the TQGs, policymakers could develop recruitment policies to *attract* high-quality teachers to disadvantaged schools. But if patterns in teacher attrition drive the observed inequities, policymakers may wish to focus on retention policies designed to *keep* high-quality teachers in disadvantaged schools. Either way, the evidence in this paper suggests that U.S. public schools have a long way to go in terms of ensuring equal access to quality teaching for advantaged and disadvantaged students.

REFERENCES

- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014a). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review*, 104(9), 2593–2632.
- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014b). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, 104(9), 2633–2679.
- Chetty, R., Hendren, N., Kline, P., & Saez, E. (2014c). Where is the land of opportunity? The geography of intergenerational mobility in the United States (No. w19843). National Bureau of Economic Research.
- Clotfelter, C. T., Ladd, H. F., & Vigdor, J. (2005). Who teaches whom? Race and the distribution of novice teachers. *Economics of Education review*, 24(4), 377-392.
- Clotfelter, C., Glennie, E., Ladd, H., & Vigdor, J. (2008). Would higher salaries keep teachers in high-poverty schools? Evidence from a policy intervention in North Carolina. *Journal of Public Economics*, 92(5), 1352-1370.
- Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2009). The academic achievement gap in grades 3 to 8. *The Review of Economics and Statistics*, 91(2), 398-419.
- Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2011). Teacher mobility, school segregation, and pay-based policies to level the playing field. *Education Finance and Policy*, 6(3), 399-438.
- Darling-Hammond, L. (2004). Inequality and the right to learn: Access to qualified teachers in California's public schools. *The Teachers College Record*, 106(10), 1936-1966.

Engel, M., Jacob, B. A., & Curran, F. C. (2013). New evidence on teacher labor supply. *American Educational Research Journal*, 0002831213503031.

Glazerman, S., & Max, J. (2011). Do low-income students have equal access to the highest-performing teachers? Mathematica Policy Research.

Goldhaber, D., Gross, B., & Player, D. (2011). Teacher career paths, teacher quality, and persistence in the classroom: Are public schools keeping their best? *Journal of Policy Analysis and Management*, 30(1), 57-87.

Goldhaber, D., Gratz, T., and Theobald, R. (2016a). What's in a teacher test? Assessing the relationship between teacher licensure test scores and student secondary STEM achievement. CALDER Working Paper 158.

Goldhaber, D., & Hansen, M. (2013). Is it just a bad class? Assessing the long-term stability of estimated teacher performance. *Economica*, 80(319), 589-612.

Goldhaber, D., Lavery, L., and Theobald, R. (2015). Uneven playing field? Assessing the teacher quality gap between advantaged and disadvantaged students. *Educational Researcher*, 44(5), 293-307.

Goldhaber, D., Lavery, L., and Theobald, R. (2016b). Inconvenient truth? Do CBAs help explain the mobility of teachers within school districts? *Journal of Policy Analysis and Management*. doi:10.1002/pam.21914

Goldhaber, D., Quince, V., and Theobald, R. (2016c). Reconciling different estimates of teacher quality gaps based on value-added. Washington, DC: National Center for the Analysis of Longitudinal Data in Education Research (CALDER), American Institutes for Research.

- Goldhaber, D. and Starz, R. (forthcoming). On the Distribution of Worker Productivity: The Case of Teacher Effectiveness and Student Achievement. *Statistics and Public Policy*.
- Goldhaber, D., Walch, J., & Gabele, B. (2014). Does the model matter? Exploring the relationship between different student achievement-based teacher assessments. *Statistics and Public Policy*, 1(1), 28-39.
- Grissom, J. A., Kalogrides, D., & Loeb, S. (2015). The micropolitics of educational inequality: The case of teacher–student assignments. *Peabody Journal of Education*, 90(5), 601-614.
- Hanushek, E. A. (1992). The trade-off between child quantity and quality. *Journal of Political Economy*, 84-117.
- Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2004). Why public schools lose teachers. *Journal of Human Resources*, 39(2), 326-354.
- Hedges, L. V., & Nowell, A. (1998). Black-white test score convergence since 1965. *The Black-White test score gap*, 149-181.
- Isenberg, E., Max, J., Gleason, P., Potamites, L., Santillano, R., Hock, H., & Hansen, M. (2013). Access to Effective Teaching for Disadvantaged Students. NCEE 2014-4001. Institute of Education Sciences.
- Isenberg, E., Max, J., Gleason, P., Johnson, M., Deutsch, J., & Hansen, M. (2016). Do low-income students have equal access to effective teachers? Evidence from 26 districts. NCEE 2017-4007. Institute of Education Sciences.

- Kalogrides, D., & Loeb, S. (2013). Different teachers, different peers: The magnitude of student sorting within schools. *Educational Researcher*, 42(6), 304-316.
- Kalogrides, D., Loeb, S., & Béteille, T. (2013). Systematic sorting: Teacher characteristics and class assignments. *Sociology of Education*, 86(2), 103-123.
- Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013). Have We Identified Effective Teachers? Seattle, WA: Bill and Melinda Gates Foundation.
- Kraft, M. A., Marinell, W. H., & Yee, D. (2015). School Organizational Contexts, Teacher Turnover, and Student Achievement: Evidence from Panel Data. *Society for Research on Educational Effectiveness*.
- Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: A descriptive analysis. *Educational Evaluation and Policy Analysis*, 24(1), 37-62.
- Lee, V. E., & Burkam, D. T. (2002). Inequality at the starting gate: Social background differences in achievement as children begin school. Economic Policy Institute.
- Mansfield, R. K. (2015). Teacher quality and student inequality. *Journal of Labor Economics*, 33(3 Part 1), 751-788.
- Owens, A., Reardon, S. F., & Jencks, C. (2016). Income segregation between schools and school districts. *American Educational Research Journal*, 0002831216652722.
- Player, D. (2010). Nonmonetary compensation in the public teacher labor market. *Education Finance and Policy*, 5(1), 82-103.
- Reardon, S. F. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. *Whither Opportunity*, 91-116.

- Reardon, S. F., & Owens, A. (2014). 60 years after Brown: Trends and consequences of school segregation. *Annual Review of Sociology*, 40, 199-218.
- Rice, J. K. (2013). Learning from experience: Evidence on the impact and distribution of teacher experience and the implications for teacher policy. *Education Finance and Policy*, 8(3), 332–348.
- Rich, M. (2014). U.S. to focus on equity in assigning of teachers. *The New York Times*, November 10, 2014.
- Rivkin, S., Hanushek, E., & Kain, J. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417-458.
- Rockoff, J. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *The American Economic Review*, 94(2), 247-252.
- Sass, T. R., Hannaway, J., Xu, Z., Figlio, D. N., & Feng, L. (2010). Value added of teachers in high-poverty schools and lower poverty schools. *Journal of Urban Economics*, 72(2), 104-122.
- Scafidi, B., Sjoquist, D. L., & Stinebrickner, T. R. (2007). Race, poverty, and teacher mobility. *Economics of Education Review*, 26(2), 14.
- Steele, J. L., Pepper, M. J., Springer, M. G., & Lockwood, J. R. (2015). The distribution and mobility of effective teachers: Evidence from a large, urban school district. *Economics of Education Review*, 48, 86-101.
- St. John, E. (2014). Understanding the factors that influence the grouping and assignment of students to elementary classrooms. Unpublished dissertation, University of Washington.

TABLES AND FIGURES

Table 1: Cross-State Data Comparison

Panel A: Statistics from 2012-13 School Year	
<p><u>North Carolina</u></p> <ul style="list-style-type: none"> • Public school students <ul style="list-style-type: none"> ○ 1,456,020 students ○ 52% White/non-Hispanic ○ 26% Black ○ 14% Hispanic ○ 54.6% FRL • Public schools <ul style="list-style-type: none"> ○ 2530 public noncharter schools ○ 108 public charter schools • Public school districts <ul style="list-style-type: none"> ○ 119 school districts • Average district size <ul style="list-style-type: none"> ○ 12,235 students ○ 22 schools 	<p><u>Washington</u></p> <ul style="list-style-type: none"> • Public school students <ul style="list-style-type: none"> ○ 1,050,900 students ○ 59% White/non-Hispanic ○ 5% Black ○ 20% Hispanic ○ 46.1% FRL • Public schools <ul style="list-style-type: none"> ○ 2678 public noncharter schools ○ 0 public charter schools • Public school districts <ul style="list-style-type: none"> ○ 295 school districts • Average district size <ul style="list-style-type: none"> ○ 3562 students ○ 9 schools
Panel B: Measures of Teacher Quality	
<p><u>North Carolina</u></p> <ul style="list-style-type: none"> • Experience <ul style="list-style-type: none"> ○ Since 1995 • Licensure test scores <ul style="list-style-type: none"> ○ Since 2000 • Value-added estimates <ul style="list-style-type: none"> ○ Since 2000 	<p><u>Washington</u></p> <ul style="list-style-type: none"> • Experience <ul style="list-style-type: none"> ○ Since 1988 • Licensure test scores <ul style="list-style-type: none"> ○ Since 2006 • Value-added estimates <ul style="list-style-type: none"> ○ Since 2007
Panel C: Measures of Student Disadvantage (School Assignment Data)	
<p><u>North Carolina</u></p> <ul style="list-style-type: none"> • % FRL <ul style="list-style-type: none"> ○ Since 1999 • % URM <ul style="list-style-type: none"> ○ Since 1995 	<p><u>Washington</u></p> <ul style="list-style-type: none"> • % FRL <ul style="list-style-type: none"> ○ Since 2002 • % URM <ul style="list-style-type: none"> ○ Since 1988
Panel D: Measures of Student Disadvantage (Student Assignment Data)	
<p><u>North Carolina</u></p> <ul style="list-style-type: none"> • FRL <ul style="list-style-type: none"> ○ Since 1999 • URM <ul style="list-style-type: none"> ○ Since 1995 	<p><u>Washington</u></p> <ul style="list-style-type: none"> • FRL <ul style="list-style-type: none"> ○ Since 2006 • URM <ul style="list-style-type: none"> ○ Since 2006

Figure 1. Student Demographics Over Time in North Carolina and Washington

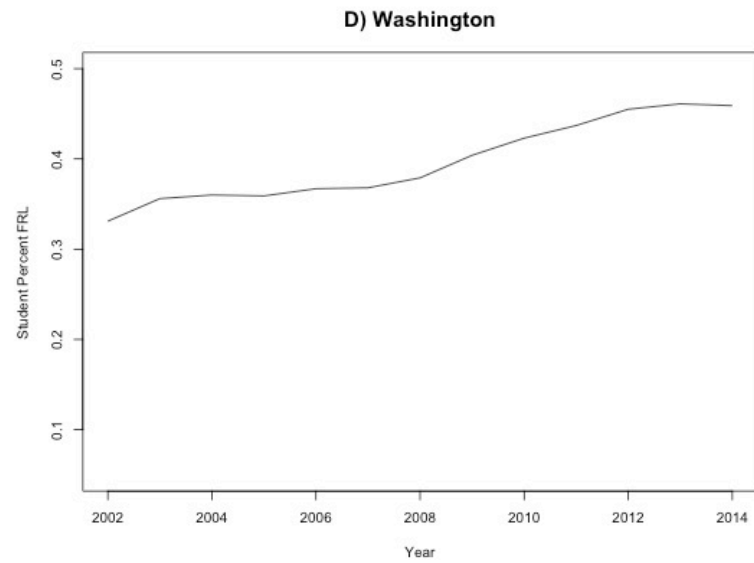
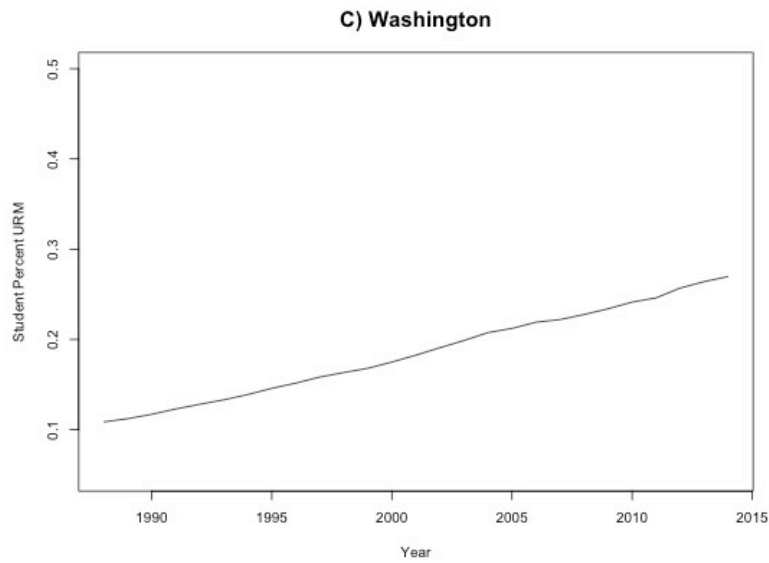
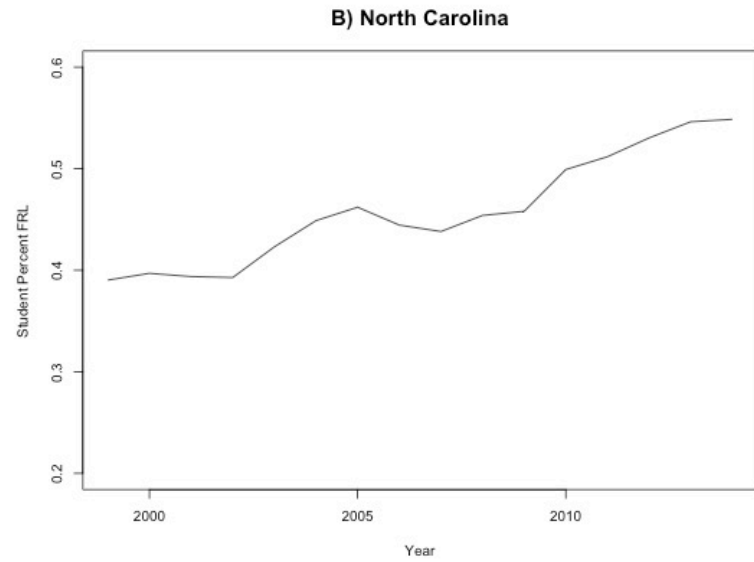
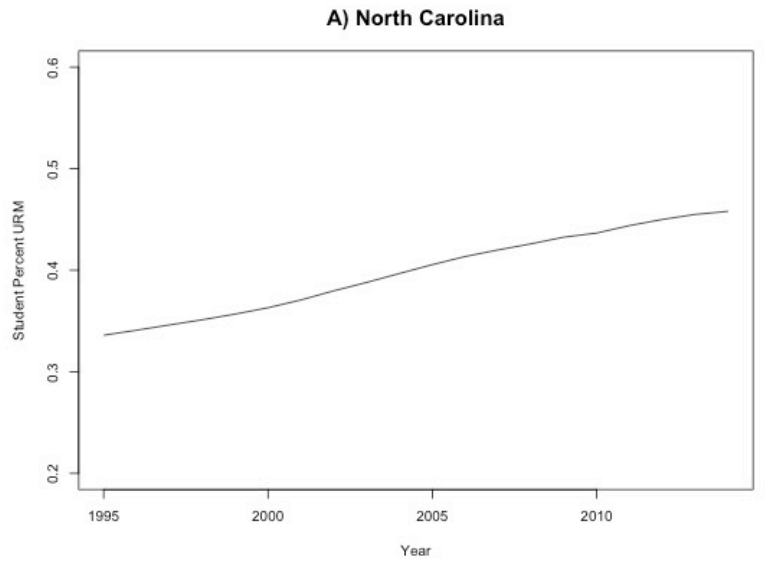


Figure 2. Geographic Distribution of Proportion of Disadvantaged Students in North Carolina and Washington In 2014

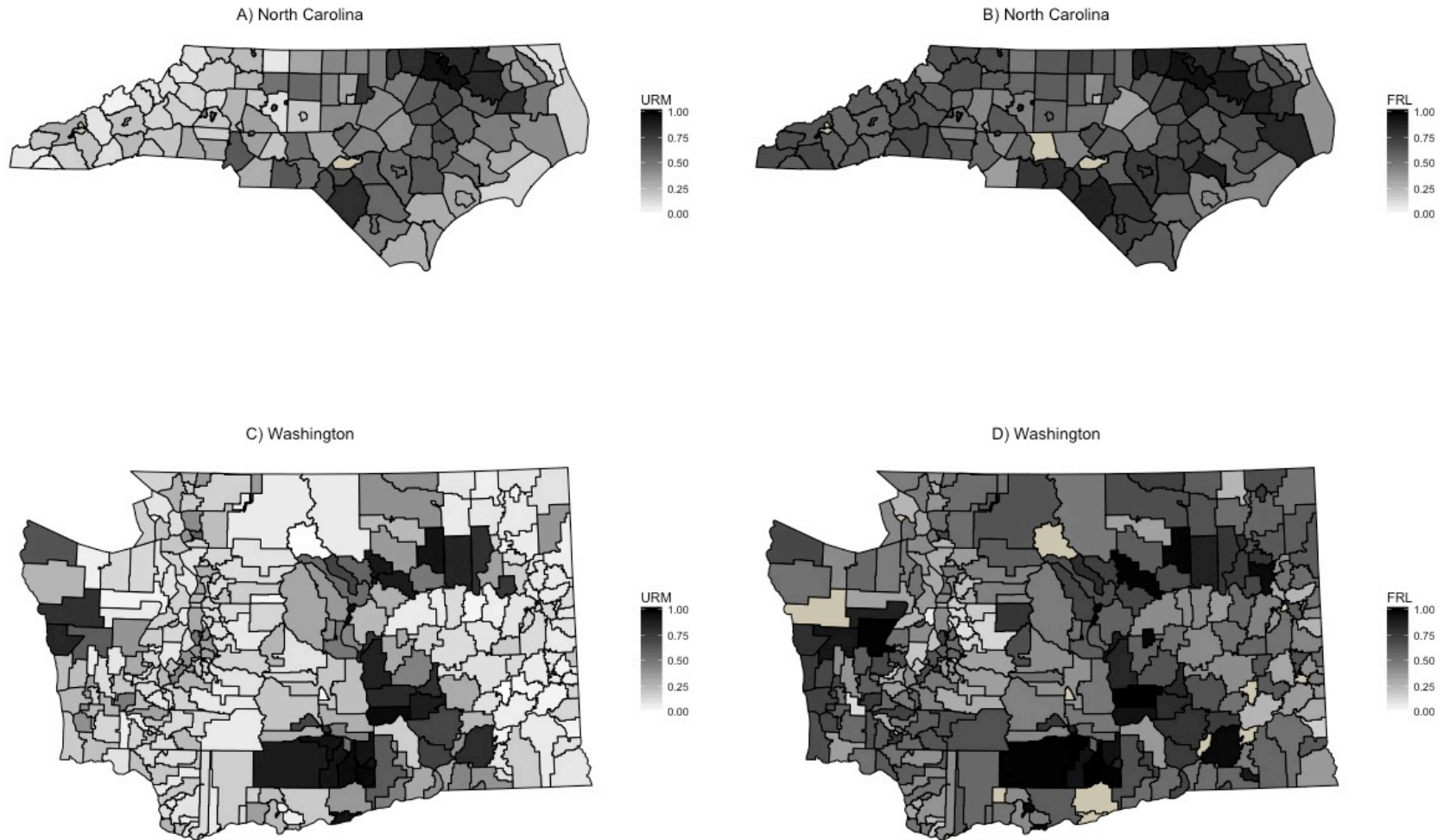


Figure 3. Segregation Indices across Districts and Schools Over Time in North Carolina and Washington

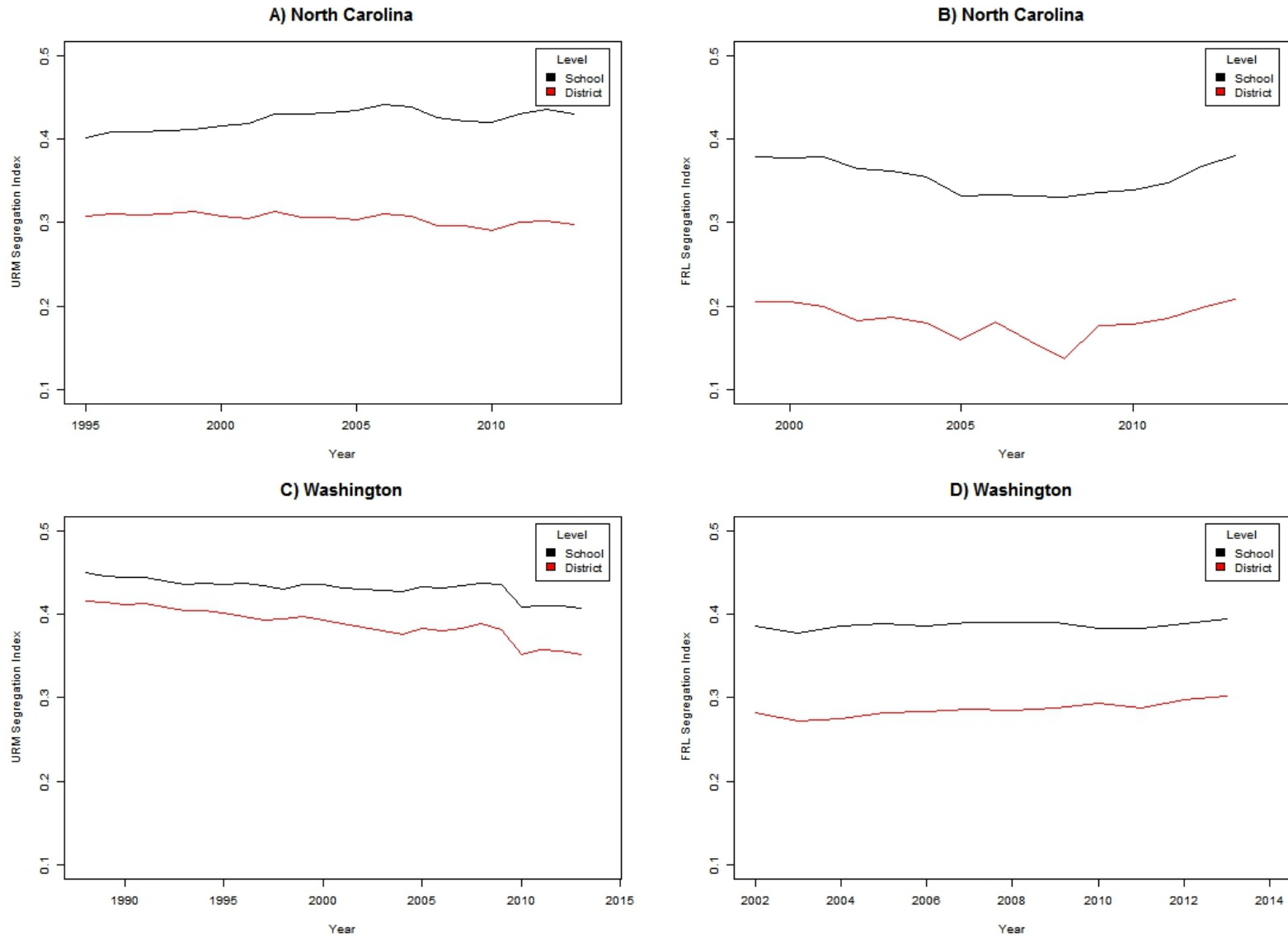


Figure 4. Exposure Rates to Novice Teachers from School Assignment Data

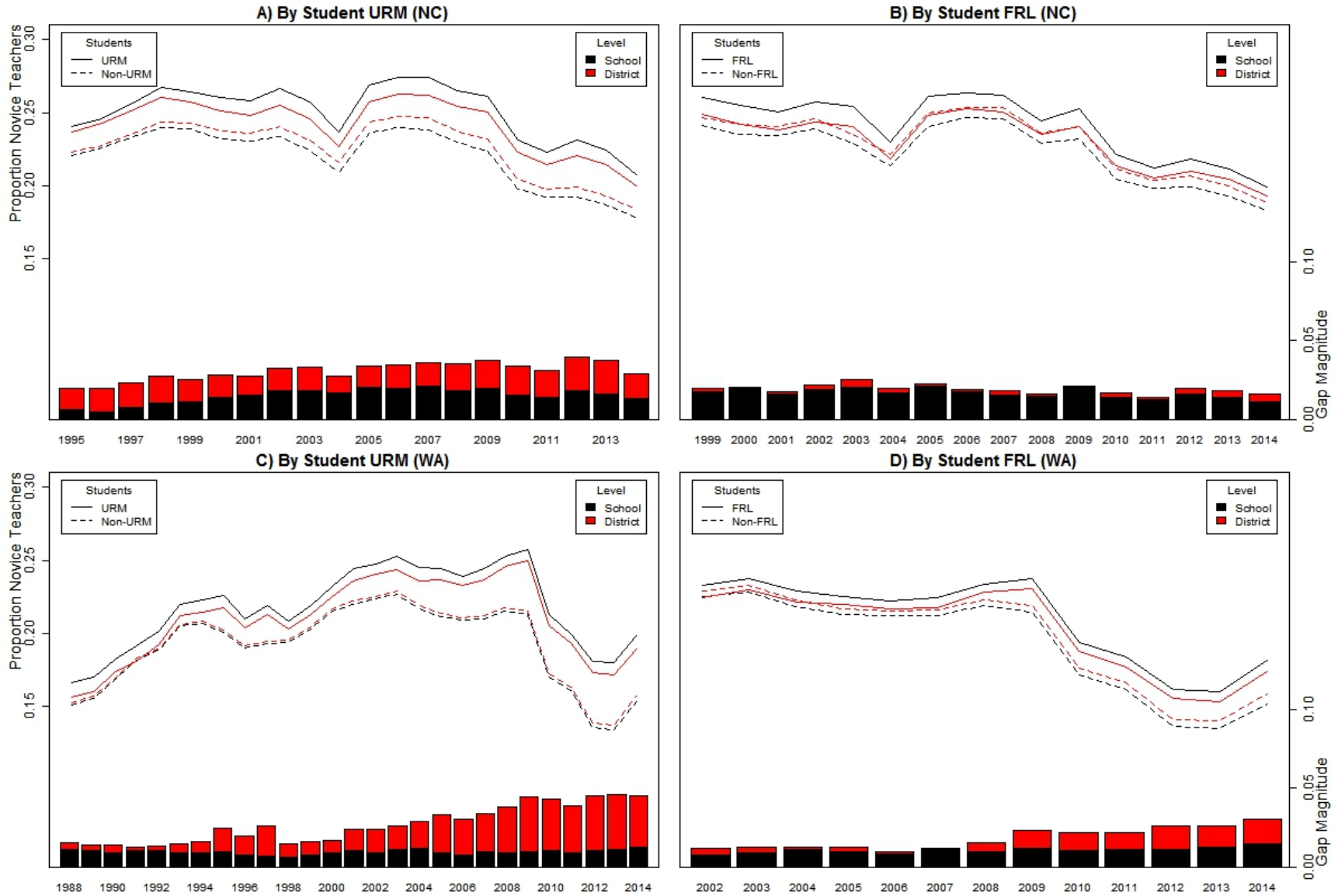


Figure 5. Exposure Rates to Teachers with Bottom-Quartile Licensure Test Score from School Assignment Data

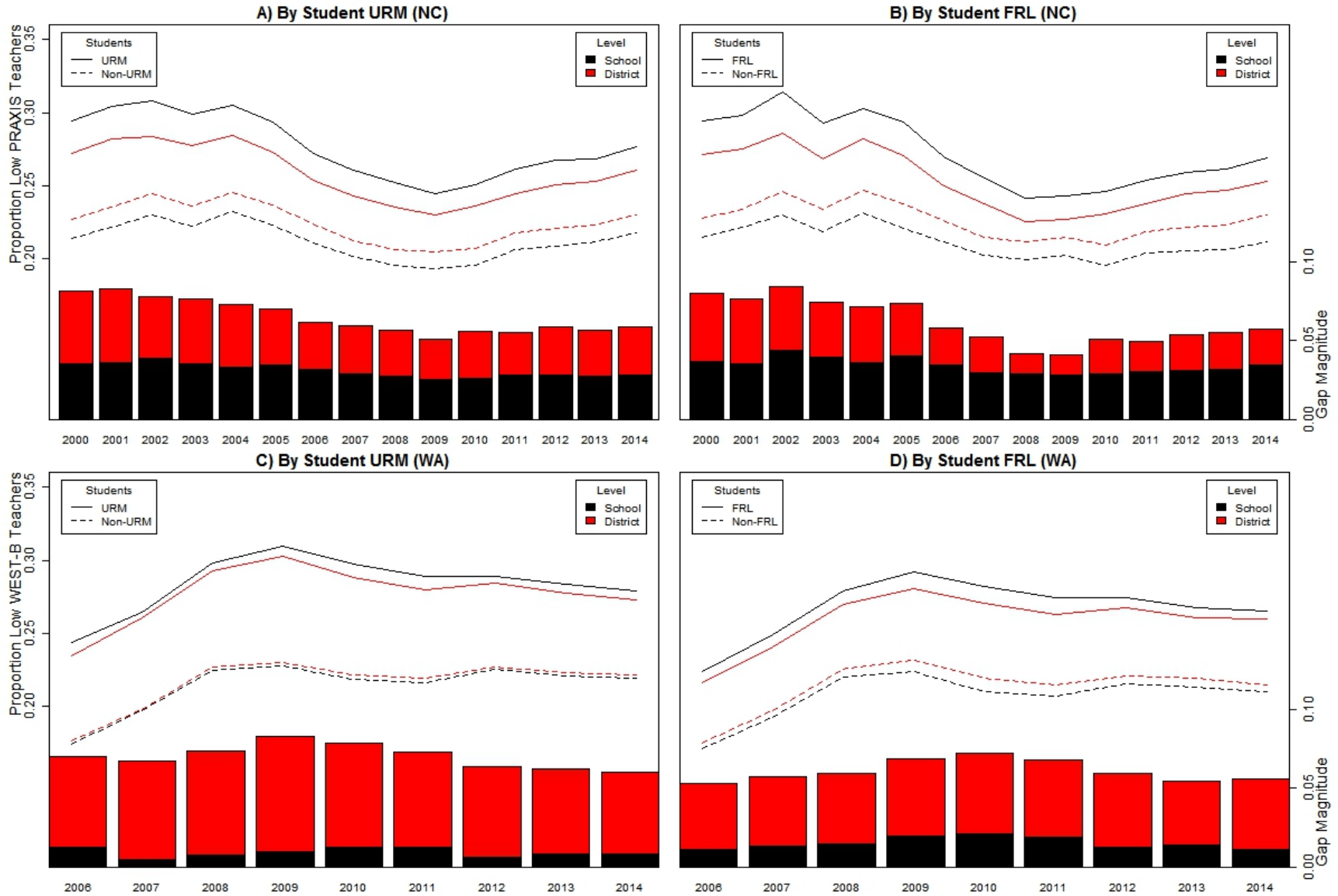


Figure 6. Exposure Rates to Teachers with Bottom-Quartile VAM Estimate from School Assignment Data

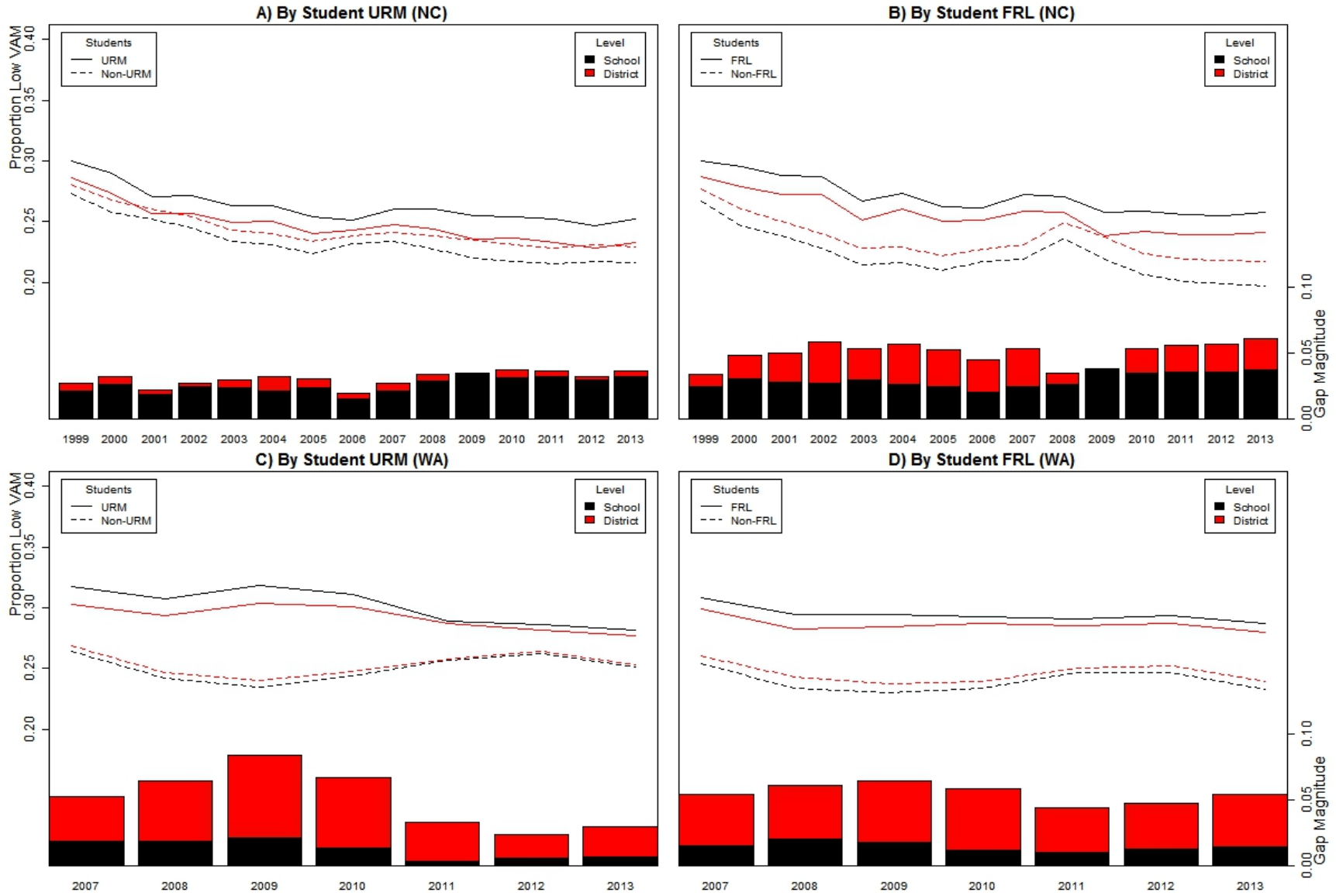


Figure 7. Exposure Rates to Novice Teachers from Student Assignment Data

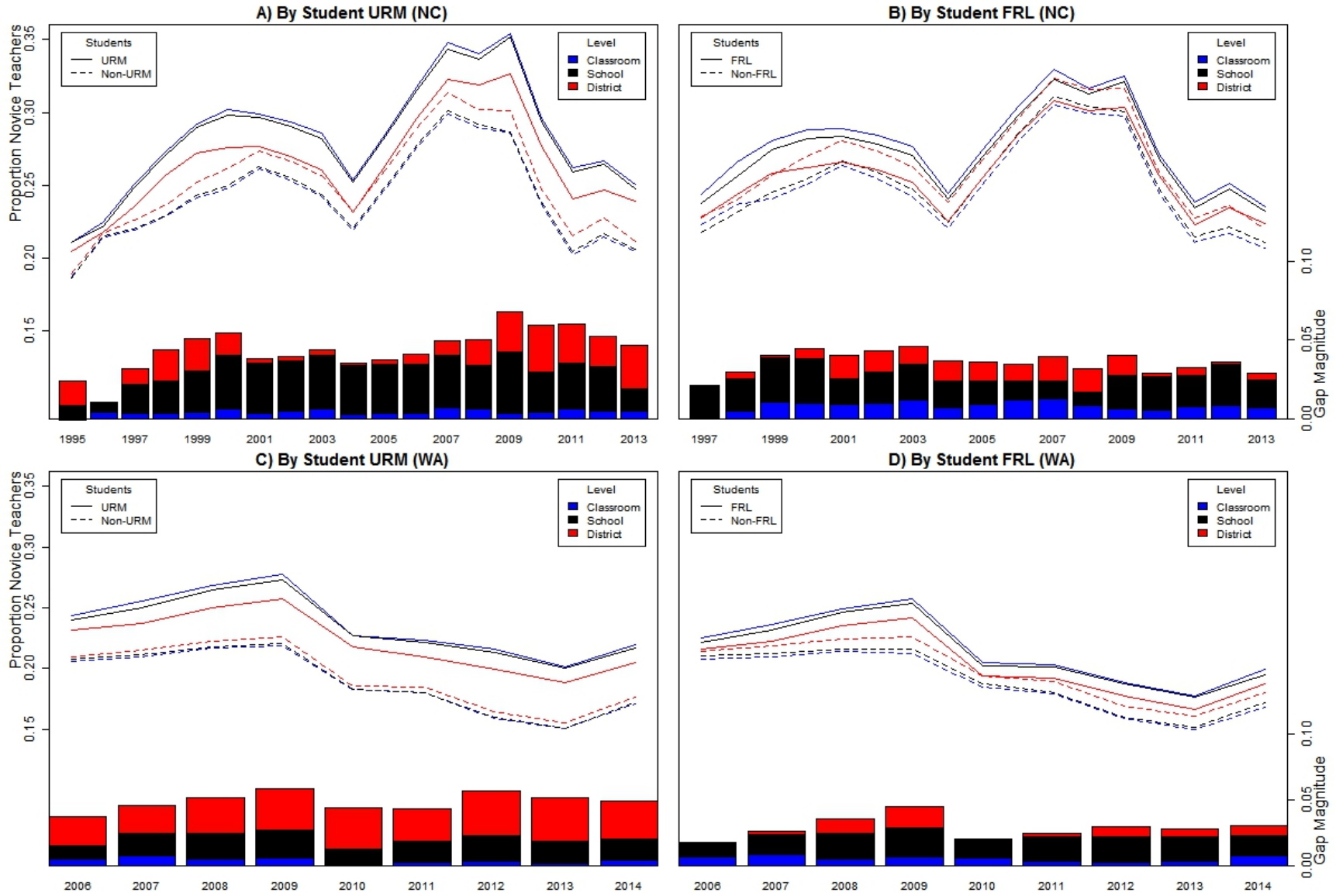


Figure 8. Exposure Rates to Teachers with Bottom-Quartile Licensure Test Score from Student Assignment Data

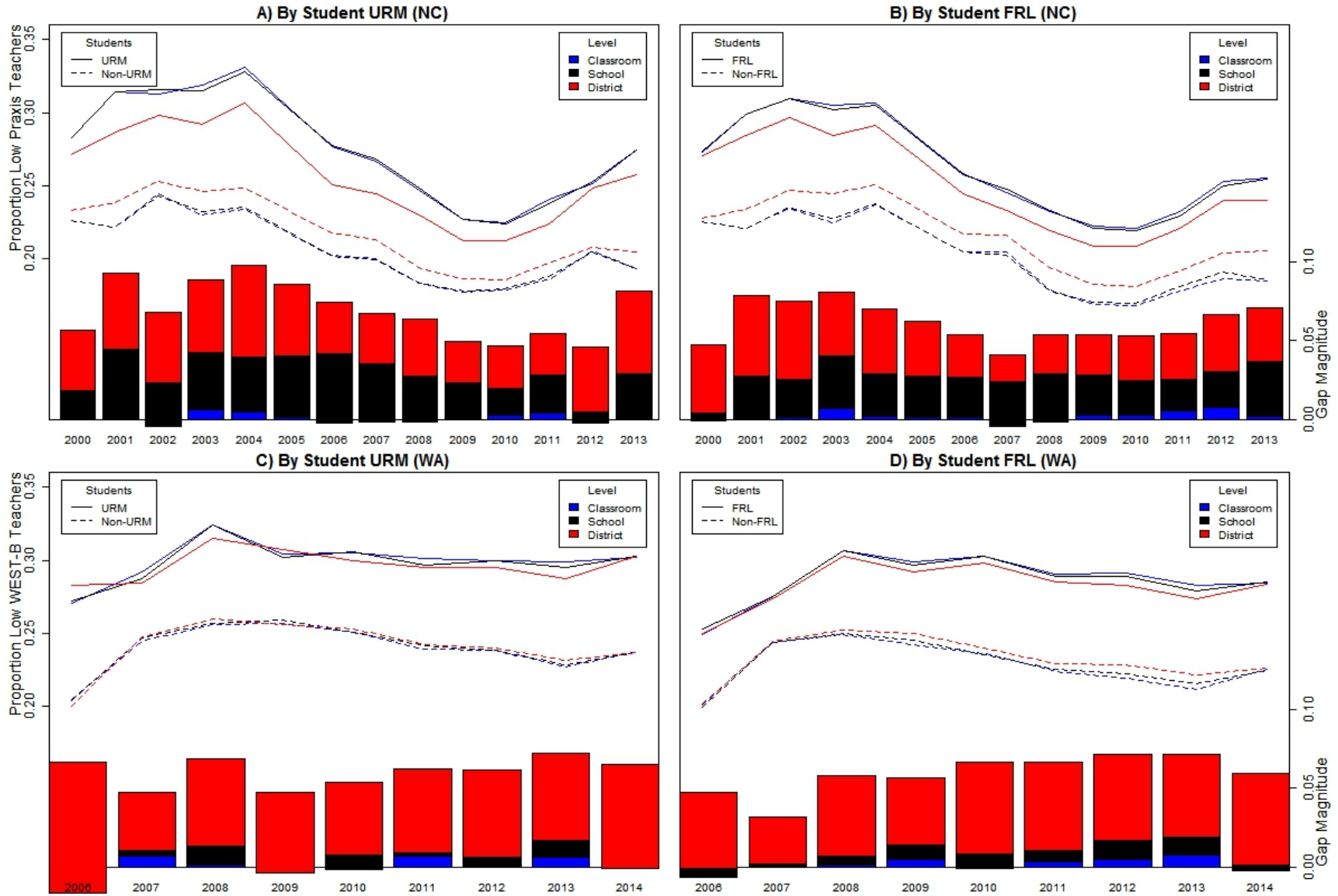


Figure 9. Exposure Rates to Teachers with Bottom-Quartile VAM Estimate from Student Assignment Data

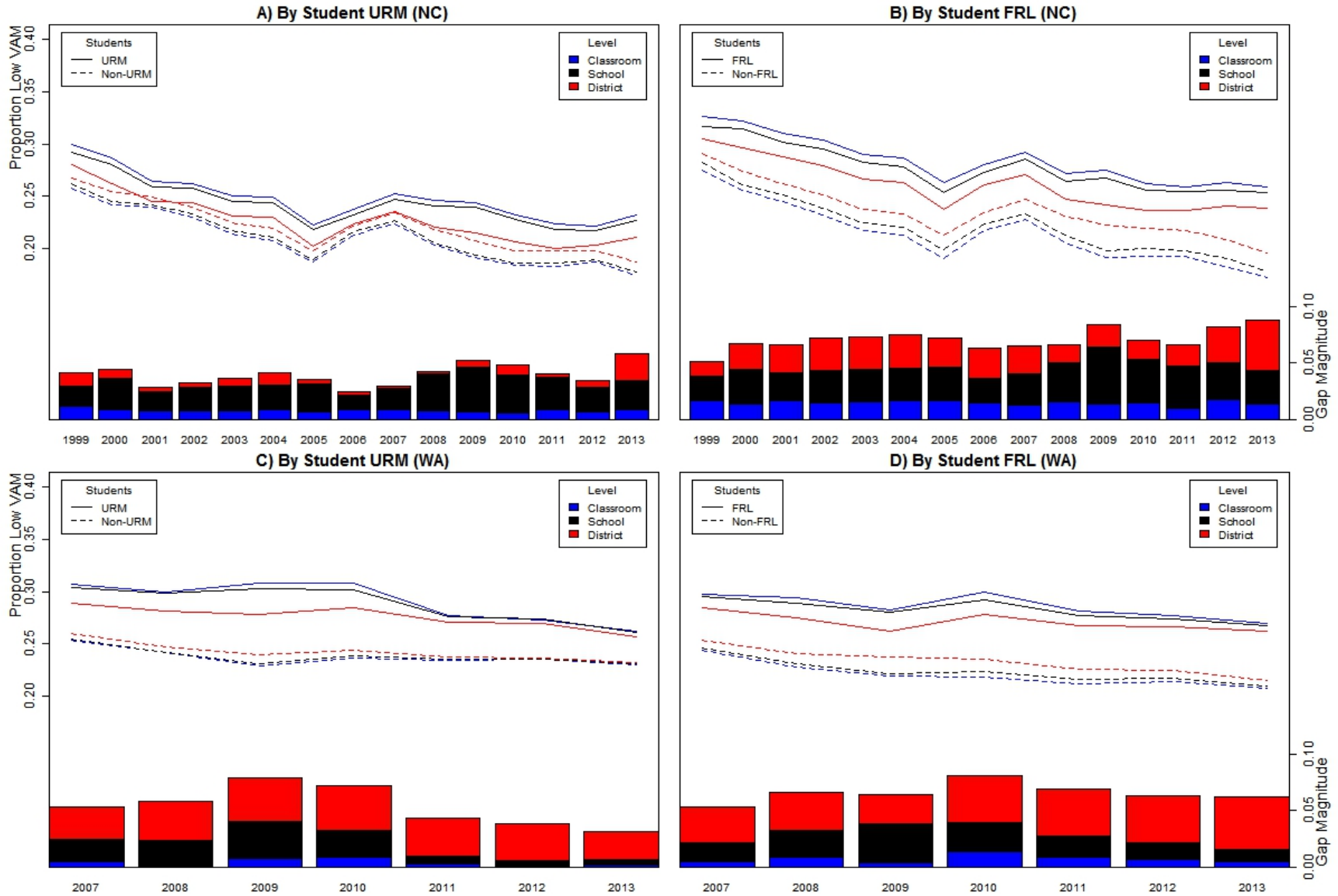


Figure 10. Estimated Cumulative Exposure to Novice Teachers in Grades 1–6, 2008-13 Cohort

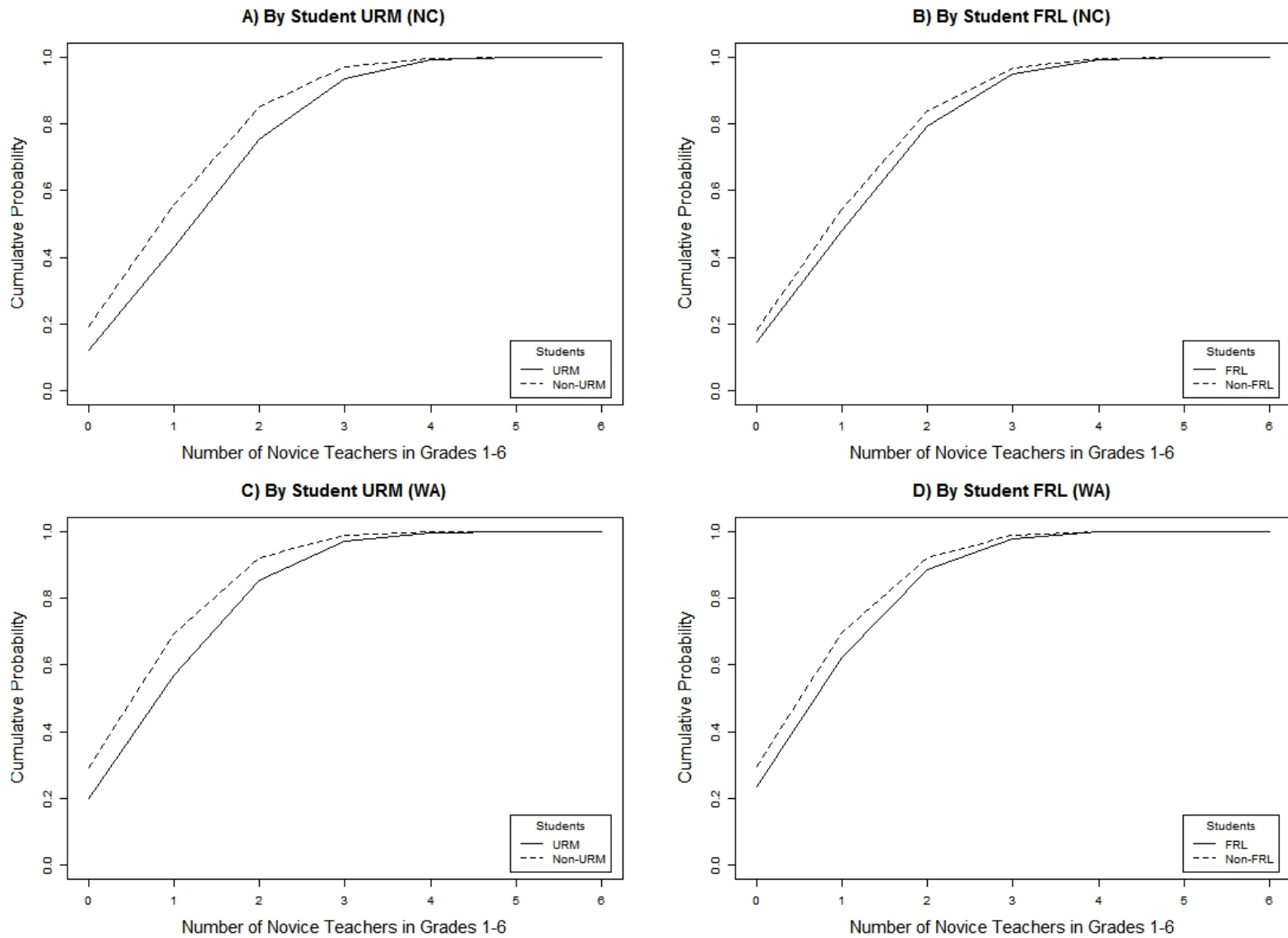


Figure 11. Estimated Cumulative Exposure to Teachers with Bottom-Quartile Licensure Test Score in Grades 1–6, 2008-13 Cohort

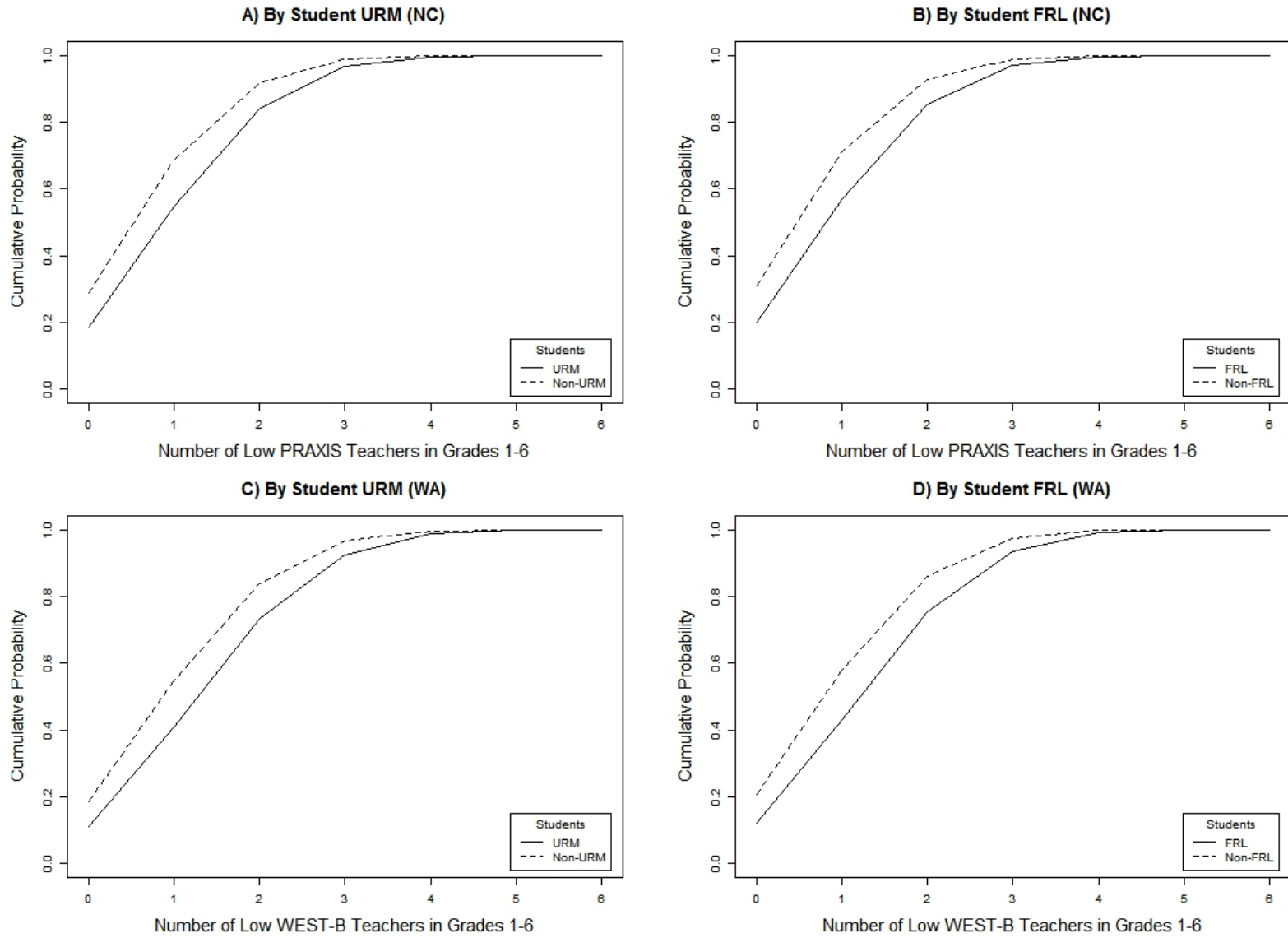


Figure 12. Estimated Cumulative Exposure to Teachers with Bottom-Quartile VAM Estimate in Grades 1–6, 2008-13 Cohort

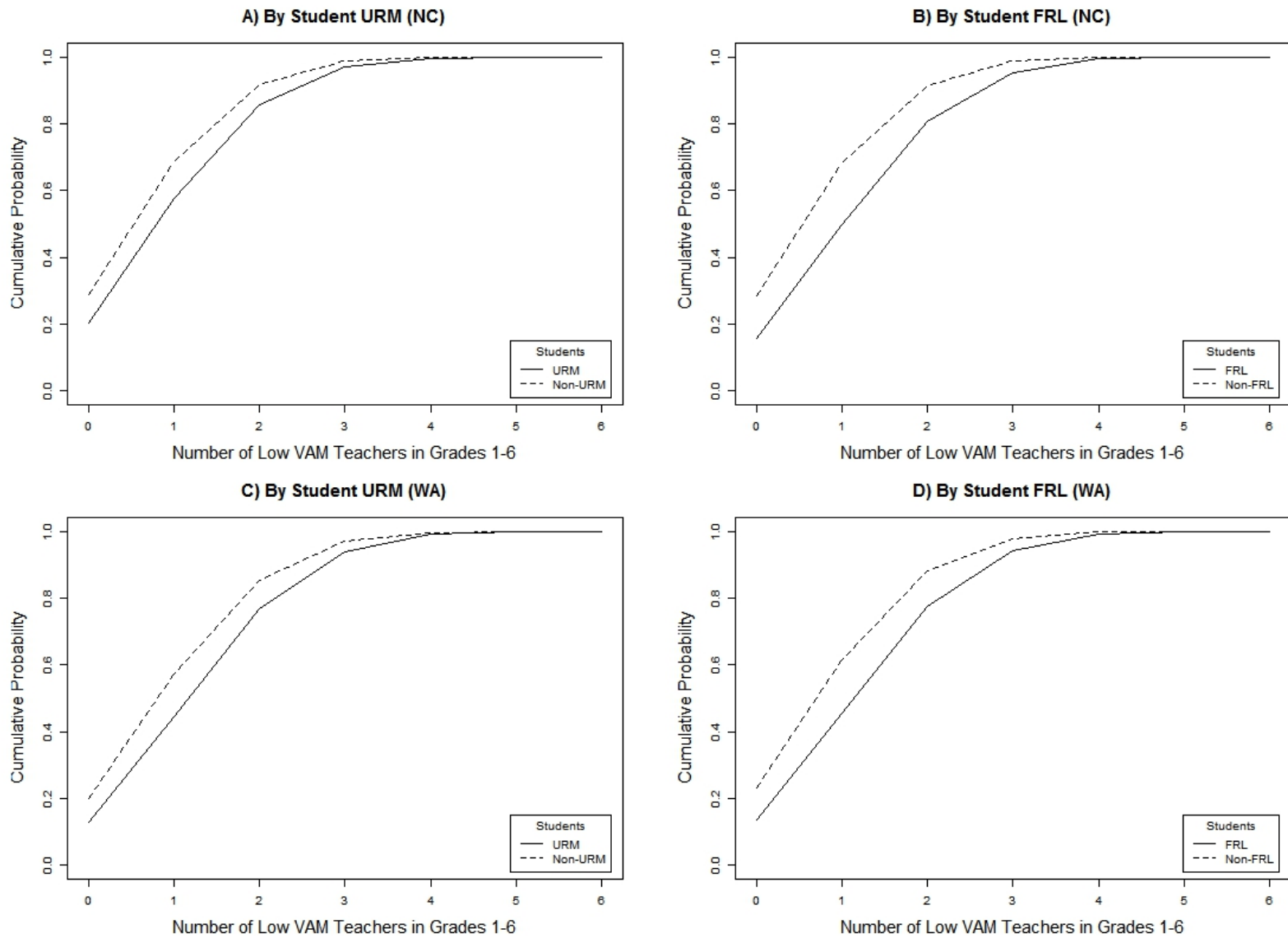


Figure 13. Geographic Distribution of TQGs in Exposure to Novice Teachers in North Carolina and Washington in 2014

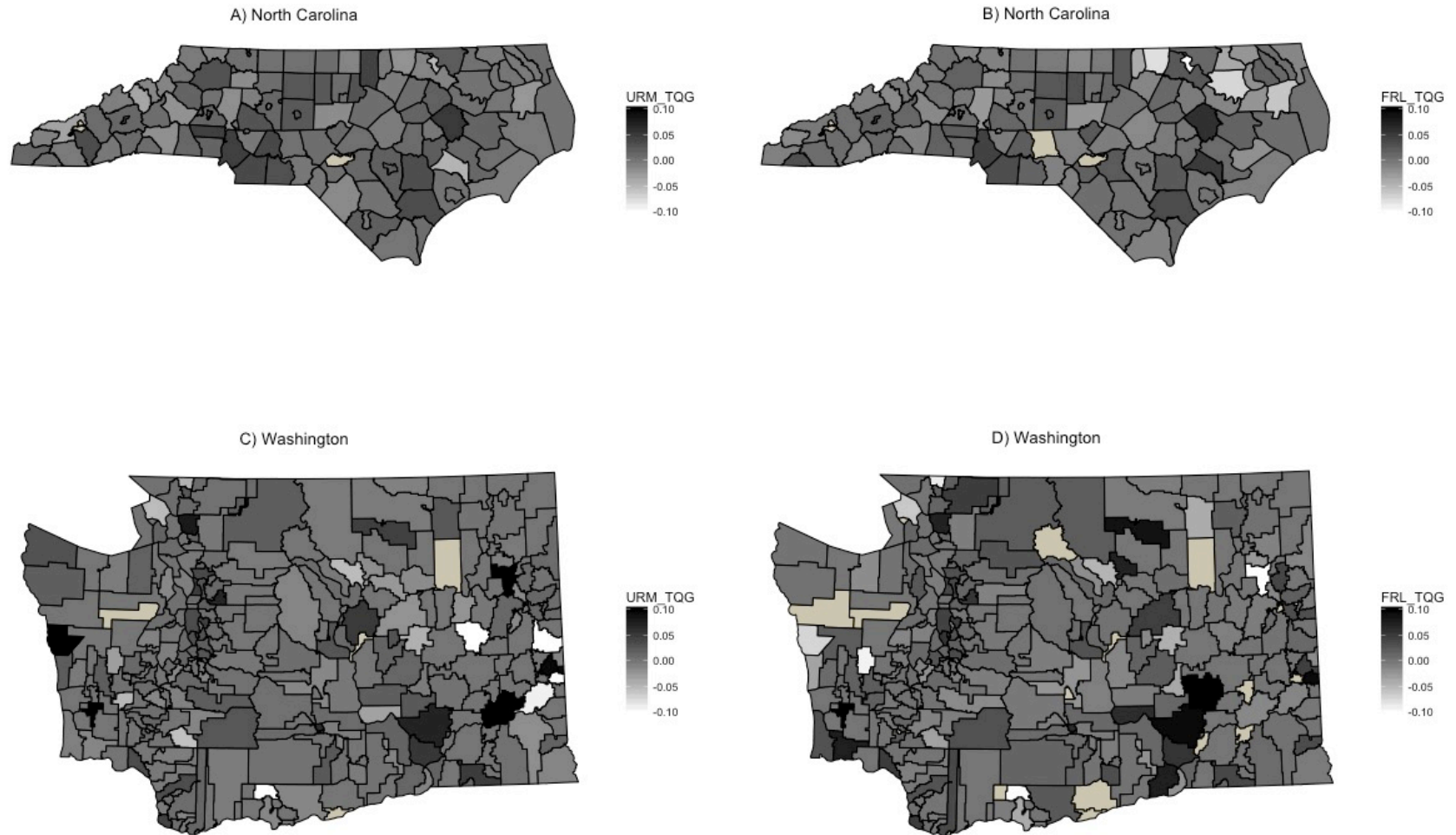


Figure 14. Average Teacher Experience from School Assignment Data

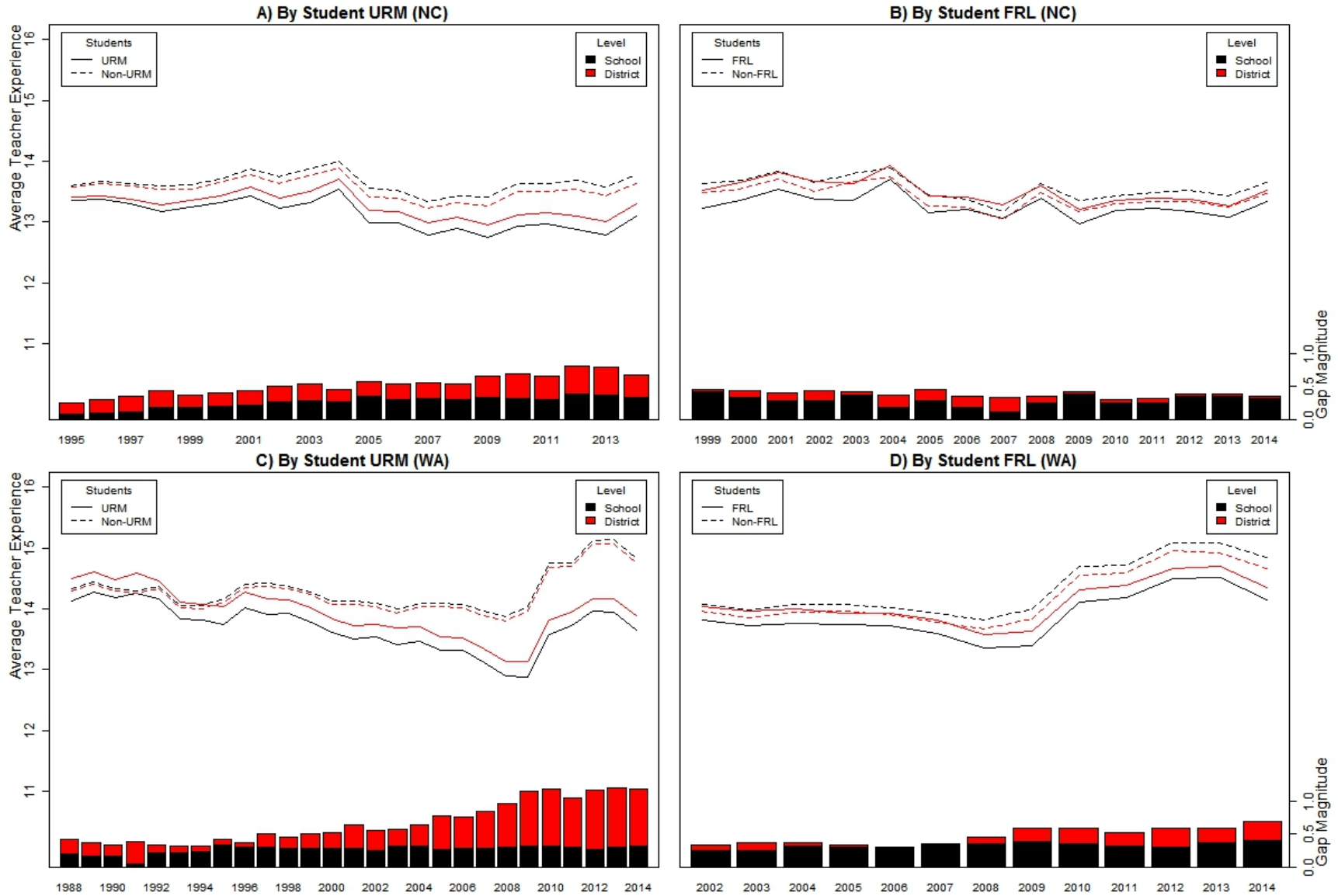


Figure 15. Average Teacher Licensure Test Scores from School Assignment Data

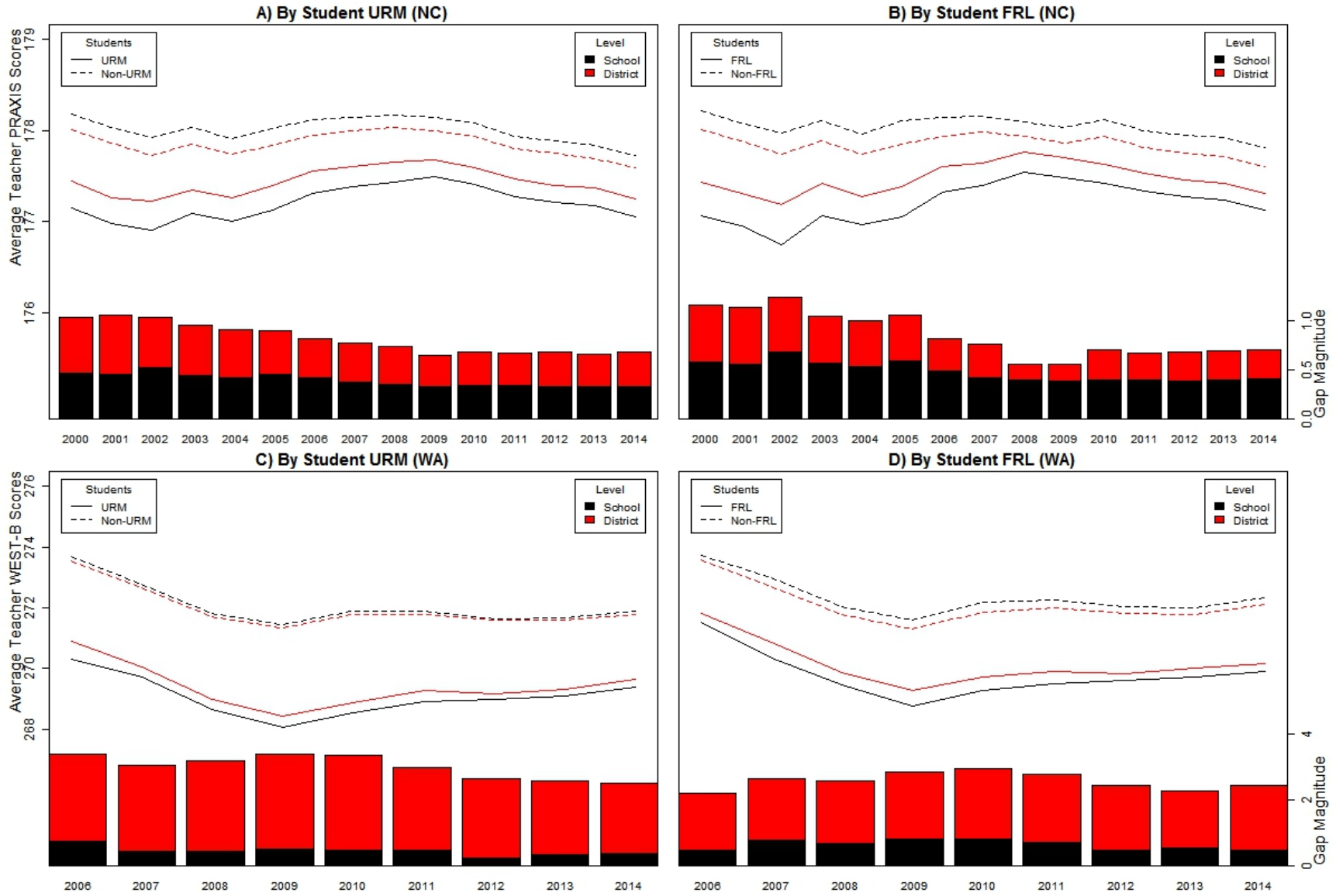


Figure 16. Average Teacher VAM Estimates from School Assignment Data

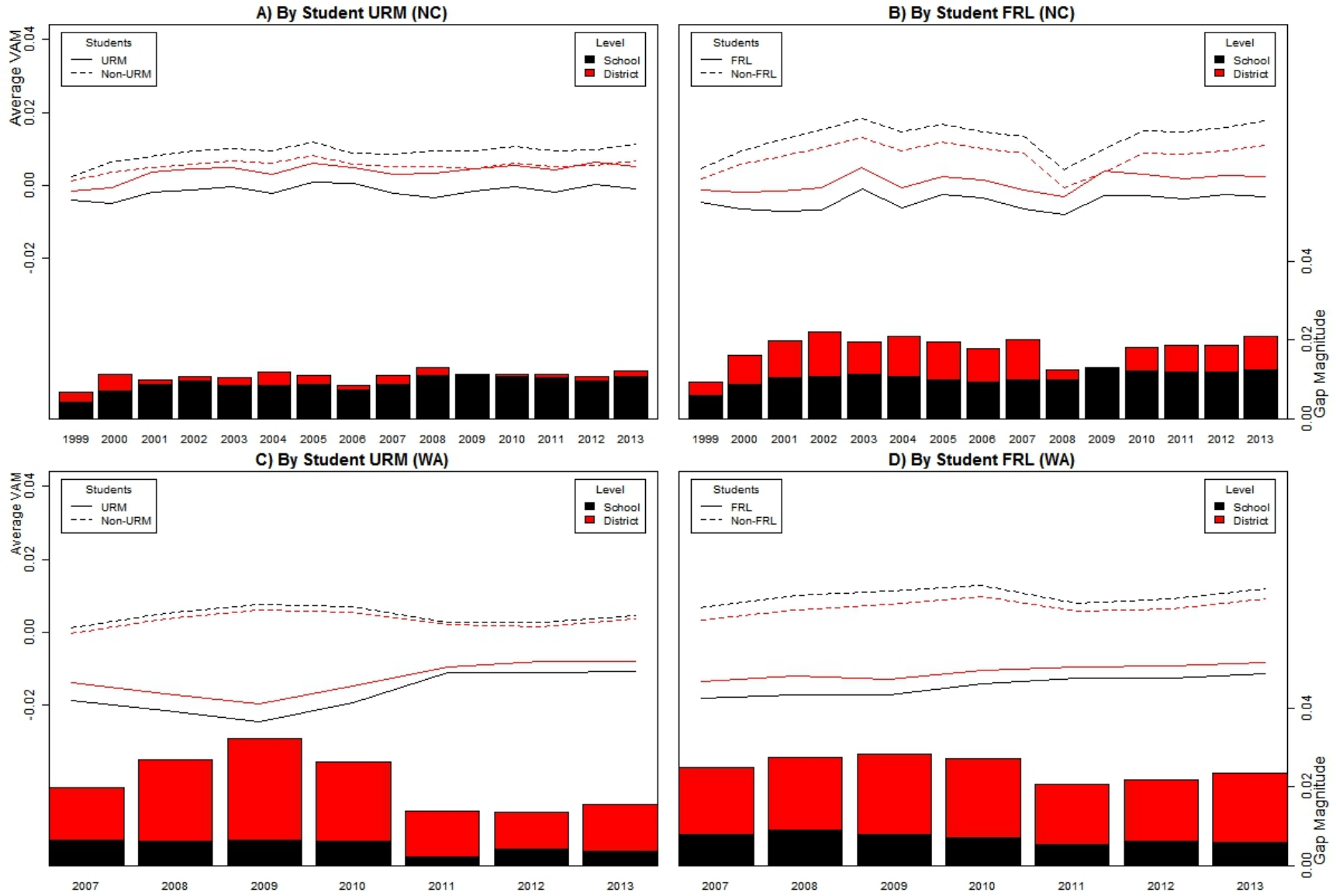


Figure 17. Average Teacher Experience from Student Assignment Data

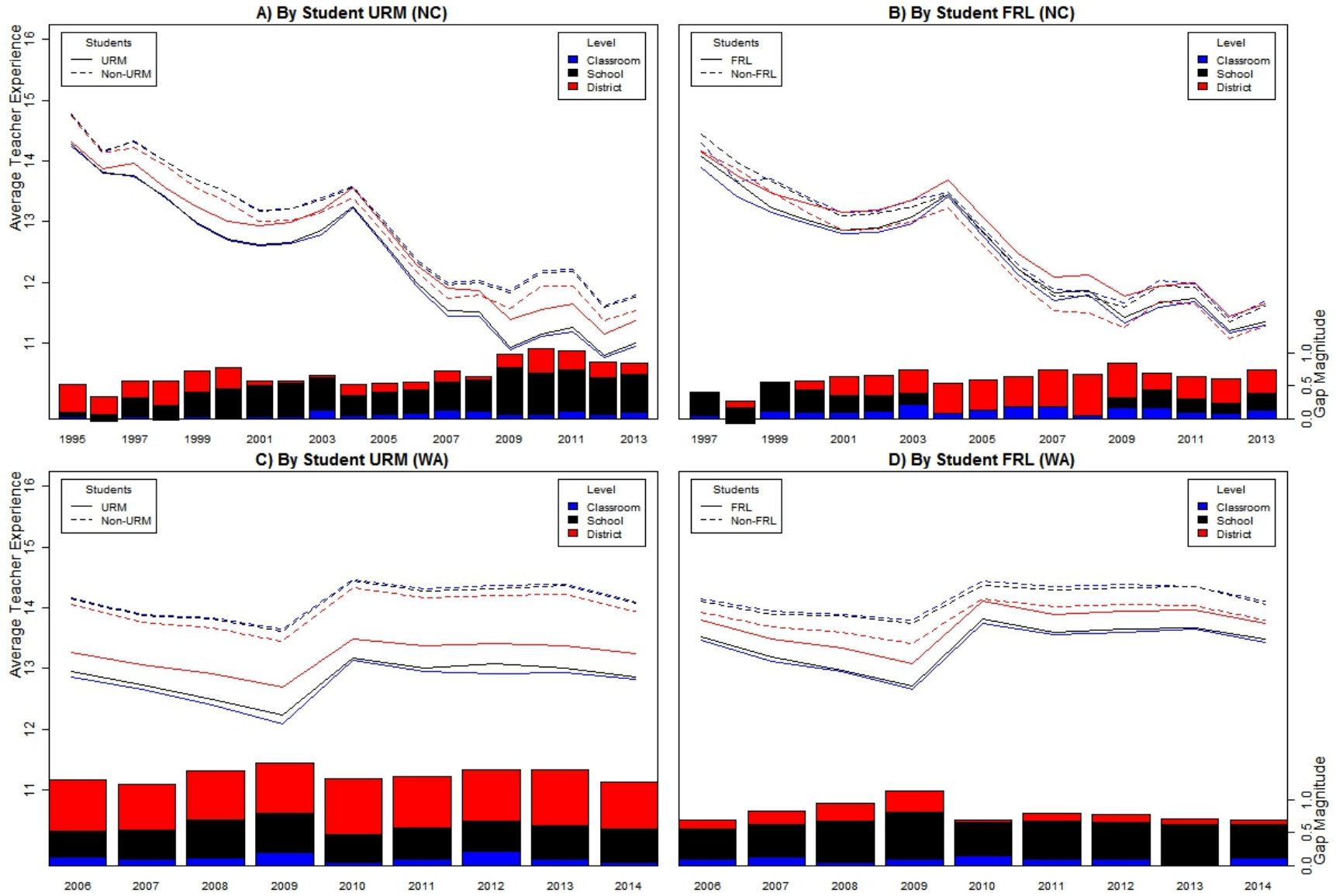


Figure 18. Average Teacher Licensure Test Scores from Student Assignment Data

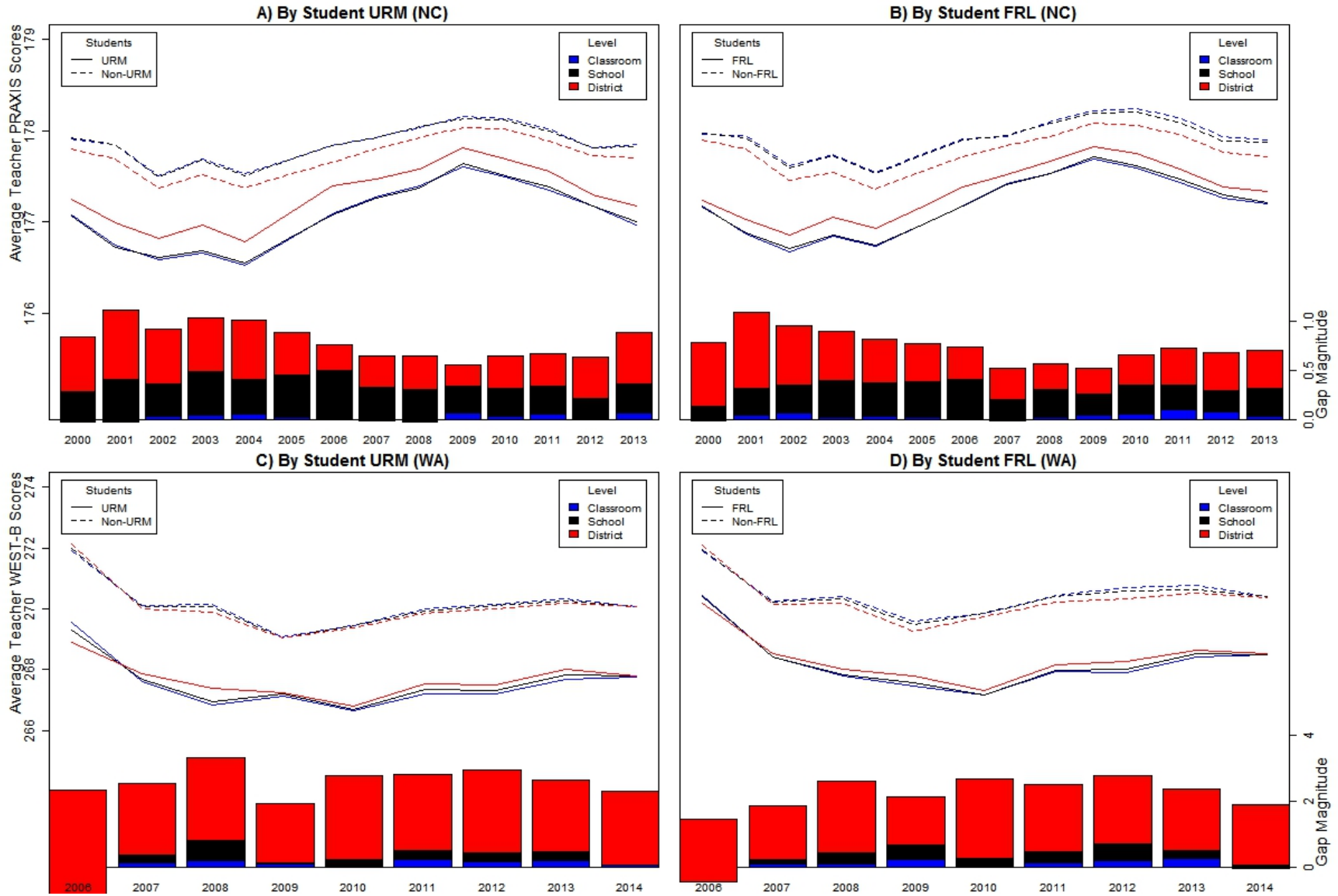


Figure 19. Average Teacher VAM Estimates from Student Assignment Data

